

THE ROLE OF CONSUMER EXPERIENCE IN THE MARKET FOR COLLEGE TEXTBOOKS

Forrest Spence

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics.

Chapel Hill
2015

Approved by:

Brian McManus

Gary Biglaiser

Donna Gilleskie

James Roberts

Helen Tauchen

© 2015
Forrest Spence
ALL RIGHTS RESERVED

ABSTRACT

FORREST SPENCE: The Role of Consumer Experience in the Market for College Textbooks.
(Under the direction of Brian McManus)

My dissertation work investigates the relationship between consumer experience in a market and decision-making. The research is unique because of the focus on learning about aggregate uncertainty as opposed to the large amount of research that has been done in the economics and marketing literatures investigating how individuals learn about their specific match with products. I use a novel dataset collected on consumers in the UNC textbook market to engage in two projects that comprise my dissertation research. The first chapter relies on a structural approach to identifying changes in consumers' search costs as they repeatedly participate in a market and quantify the welfare implications that result from differences in market tenure. The second chapter investigates the extent to which consumers' beliefs about price distributions reflect actual empirical distributions and whether these beliefs become more accurate with experience.

For my parents and their contagious love of learning.

ACKNOWLEDGMENTS

I am very grateful for the guidance I have received from the five dissertation committee members. First and foremost, I would like to thank my dissertation advisor, Brian McManus, for the countless hours he has spent mentoring me. My growth professionally during this process is largely attributable to him. I would like to particularly thank him for encouraging me to express my opinions and being patient with me when we did not see eye to eye. I thank Gary Biglaiser for encouraging me to keep my work grounded in economic theory and for always being supportive. I thank Donna Gilleskie for being such an incredible role model both professionally and personally. Her support and mentorship have been an inspiration to myself and countless other students. I thank James Roberts for always pushing me to be better. While I will never be able to keep up with him intellectually, I appreciate his humor and patience as I have tried. I thank Helen Tauchen for her guidance during the entire doctoral process. She has been instrumental in the successful completion of my Ph.D. In addition to my committee members, I am indebted to Christopher Cronin, Brett Matsumoto, Tiago Pires, and Uyen Tran for their help and constructive criticism. I am also grateful to Patrick Conway, Neville Francis, and Rita Balaban for the important roles they have played in my successful completion. Finally, I would like to thank Jeremy Petranka for his guidance. I would not be half as effective of a teacher that I am today without the tools and inspiration he has given to me.

TABLE OF CONTENTS

| | |
|---|-----------|
| LIST OF TABLES | ix |
| LIST OF FIGURES | xi |
| 1 Introduction | 1 |
| 2 Experience and the Value of Search | 3 |
| 2.1 Introduction | 3 |
| 2.2 Literature Review | 8 |
| 2.2.1 Constrained Choice Sets | 8 |
| 2.2.2 Learning | 11 |
| 2.3 Empirical Model | 12 |
| 2.3.1 Full Choice Set | 12 |
| 2.3.2 Limited Choice Set | 13 |
| 2.3.3 Search Decision | 14 |
| 2.3.4 Alternative-Specific Utility | 17 |
| 2.3.5 Reservation Utility Rule | 20 |
| 2.3.6 Likelihood Function | 22 |
| 2.4 Data | 24 |
| 2.4.1 Online Questionnaire Data | 24 |
| 2.5 Results | 28 |
| 2.5.1 Parameter Estimates | 28 |

| | | |
|----------|--|-----------|
| 2.5.2 | Elasticities | 29 |
| 2.5.3 | Counterfactuals and Model Fit | 30 |
| 2.6 | Conclusion | 31 |
| 3 | Price Beliefs and Experience | 37 |
| 3.1 | Introduction | 37 |
| 3.2 | Theoretical Motivation | 39 |
| 3.3 | Data | 41 |
| 3.3.1 | Textbook Purchasing Scenarios | 41 |
| 3.3.2 | Online Questionnaire Data | 43 |
| 3.3.3 | Online Retailer Data | 44 |
| 3.4 | Results | 46 |
| 3.4.1 | Expectations Results | 47 |
| 3.4.2 | Distribution Results | 50 |
| 3.4.3 | Price Beliefs by Major | 54 |
| 3.5 | Learning vs. Selection | 55 |
| 3.6 | Conclusion | 56 |
| A | Appendix to Chapter 2: Experience and the Value of Search | 67 |
| A.1 | Estimation | 67 |
| A.1.1 | Structural Search Model Estimation | 67 |
| A.1.2 | Consideration Set Approach Estimation | 74 |
| A.2 | Data | 84 |
| A.2.1 | Price Issues | 84 |
| A.2.2 | Online Questionnaire | 85 |
| A | Appendix to Chapter 3: Price Beliefs and Experience | 99 |
| A.1 | Data Appendix | 99 |

| | | |
|---------------------|--|------------|
| A.1.1 | Online Questionnaire Data | 99 |
| A.1.2 | Textbook Purchasing Scenarios | 104 |
| A.1.3 | Survey Attrition and Estimation Sample | 106 |
| A.2 | Robustness Checks | 106 |
| A.2.1 | Omission Criteria | 107 |
| A.2.2 | Price Distribution | 110 |
| BIBLIOGRAPHY | | 114 |

LIST OF TABLES

| | | |
|------|---|----|
| 2.1 | Semesters Enrolled | 33 |
| 2.2 | Student Demographics | 33 |
| 2.3 | Students' Beginning of Semester Textbook Decisions | 33 |
| 2.4 | Students' Retailer Choices | 33 |
| 2.5 | Students' Online Retailer Choices - Purchased or Rented | 34 |
| 2.6 | Students' Purchasing Behavior | 34 |
| 2.7 | Students' Stated Preferences | 34 |
| 2.8 | Correlation in Students' Stated Preferences | 34 |
| 2.9 | Distribution of Payments for Textbooks | 35 |
| 2.10 | Search Cost Estimates | 35 |
| 2.11 | Utility Parameter Estimates | 35 |
| 2.12 | Elasticity Estimates | 36 |
| 2.13 | Eighth Semester Search Costs Counterfactual | 36 |
| 3.1 | Respondent Characteristics | 59 |
| 3.2 | Ratio of Amazon Prices to Bookstore Prices by Survey Period | 59 |
| 3.3 | Mean Ratio Comparisons by Online Purchasing Experience, Used Books | 60 |
| 3.4 | Mean Difference Comparisons by Online Purchasing Experience, Used Books | 60 |
| 3.5 | Mean Ratio Comparisons by Online Purchasing Experience, New Books | 61 |
| 3.6 | (Online Expectation / Bookstore Price) Regressed on Prev. Purchases | 62 |
| 3.7 | (Online Expectation / Bookstore Price) Regressed on Experience | 63 |
| 3.8 | Reported Probability that Lowest Price $< b * \text{Expected Lowest Price}$ | 63 |
| 3.9 | Mean and Variance Comparisons (Log-Normal Assumption) | 64 |

| | |
|---|-----|
| 3.10 Mean Comparison by Major | 65 |
| 3.11 Variance Comparison by Major (Log-Normal Assumption) | 66 |
| 3.12 Parameter Values by Change in Experience | 66 |
| A.1 : Textbook Scenarios | 105 |
| A.2 : Survey Attrition | 106 |
| A.3 : Results Using Alternative Samples | 107 |
| A.4 : Distribution Results | 109 |

LIST OF FIGURES

| | | |
|-----|--|-----|
| 3.1 | Textbook Purchasing Scenario | 58 |
| 3.2 | Used price pdf versus empirical dist. by level of experience | 61 |
| 3.3 | New price pdf versus empirical dist. by level of experience New_{alt} is the minimum price for a new textbook from Amazon or Amazon Marketplace. | 65 |
| A.1 | Histograms of Prices | 111 |
| A.2 | Kernel Density Estimate | 111 |
| A.3 | Kernel Density Estimate | 112 |
| A.4 | Daily Used Prices | 112 |
| A.5 | Daily New Prices | 113 |

CHAPTER 1

INTRODUCTION

My dissertation work investigates the relationship between consumer experience in a market and decision-making. The research is unique because of the focus on learning about characteristics of a market as opposed to characteristics of a product itself. A significant amount of research has been done in the economics and marketing literatures to understand how individuals learn about their specific match with products and the implications this has for consumer decision-making (e.g., brand loyalty, strategic experimentation). However, little research has been done to understand how consumers learn about characteristics of a market that are common to all individuals. I use a novel dataset collected on consumers in the UNC textbook market to engage in three projects that comprise my dissertation research. The first paper relies on a structural approach to identifying changes in consumers' value of online search and the resulting welfare effects of experience. The second paper relies on more reduced form methods to identify the impact of experience on consumer behavior.

My first chapter, "Consumer Experience and the Value of Search in the Online Textbook Market," focuses on how consumers learn about the benefits and costs of search in the online textbook market as they gain experience in the market. I collect detailed individual level purchasing and search data on 6,000 consumers in the UNC textbook market. I use these data to estimate a demand model that incorporates limited information and costly search as a function of past market experience. Consumers' expected value of search in the online textbook market increases significantly as they gain experience in the market, even after controlling for differences in individual, course, and textbook characteristics. Demand and search estimates imply that experienced consumers have an average welfare gain of approximately 5% of the price of purchase (\$8 per textbook

assignment) due to differences in their expected value of search in the online market. This result implies that there could be significant gains to providing information about market characteristics to consumers, particularly in markets where a large portion of the consumers are inexperienced (e.g., durable goods markets).

In work with Brett Matsumoto, “Price Beliefs and Experience: Do Consumers’ Beliefs Converge to Empirical Distributions with Repeated Purchases?,” we investigate one specific mechanism in which individuals learn about market characteristics. In particular, we investigate to what extent consumer beliefs about the distribution of prices reflect the actual empirical price distribution and whether consumers learn about features of the empirical distribution through experience. Using subjective beliefs for online textbook prices from a survey of 1,200 college students, we find that consumers with no prior experience in purchasing textbooks online have beliefs that are biased upward relative to the empirical distribution. However, more experienced consumers generally have more accurate beliefs about the price distribution which provides an explanation for their higher observed levels of search in the online market.

CHAPTER 2

EXPERIENCE AND THE VALUE OF SEARCH

2.1 Introduction

First-time consumers in a market face several challenges. Information on the existence of retailers, the products they offer, and the prices and characteristics of these products are all costly to obtain. As consumers repeatedly participate in a market and interact with other market participants, they may acquire information about characteristics of the market that are common across all consumers. This paper investigates how consumers' decisions to search across retailers change as they learn about these market-level characteristics.

A considerable amount of research in the economics and marketing literatures focuses on how individuals learn about match-specific characteristics (e.g., Erdem and Keane, 1996; Akerberg, 2003), but little work has been done to understand how individuals learn about characteristics of a market as a whole. For instance, Crawford and Shum (2005) examine the importance of individuals' uncertainty about the efficacy of anti-ulcer medications (a match-specific characteristic). However, consumers may also face uncertainty about the availability of alternative medications, which retailers offer these medications, and the distribution of retailers' prices (market-level characteristics).

From a policy perspective, it is important to distinguish whether consumers are learning primarily about match-specific or market-level characteristics. Informational policies can be used to alleviate uncertainty about market-level characteristics, but will be ineffective at changing behavior in markets where the consumers primarily face uncertainty regarding match-specific characteristics. In order to evaluate the potential value of informational policies in a market, it is critical to

understand if there are systematic differences in behavior across consumers that result from learning about market-level characteristics. If these differences exist, policies that provide consumers with information about market-level characteristics can increase the rate of learning for inexperienced consumers in a market and improve consumer outcomes. This may be particularly effective in markets where a large proportion of the consumers have made relatively few purchases (e.g., durable goods or health care markets).

Experience appears to play a large role in consumer decision-making in the college textbook market. First-year students (generally first-time participants) tend to purchase their textbooks at university bookstores much more frequently than upper-year students (experienced participants). All participants in the market are generally provided with information about the campus bookstore, but may significantly differ in the amount of information they have about alternatives available outside of the campus bookstore. In particular, experienced consumers may have learned characteristics of the online market through repeated participation that changes their purchasing behavior.

As consumers gain information about the online market, this affects their decision to engage in search across retailers. The incentives to engage in search are broadly defined as consumers' value of search: consumers' beliefs about the benefits of search minus the cost of engaging in search. This paper contributes to the literatures on consumer search and learning by evaluating the level of heterogeneity in consumers' value of search, the degree to which this value varies by experience gained through repeated participation in a market, and the benefits consumers receive from increased information about characteristics of the online textbook market.

This paper makes a policy contribution by calculating the potential welfare gains that would result from eliminating the gap in consumers' expected value of search in the online market across experience levels. If experience leads to less uncertainty about the availability of online retailers, their reliability, and their prices, policies that provide information to consumers can increase the rate of learning over characteristics of the online market, thereby removing welfare gaps attributable to differences in experience. Policies that have similar effects have been enacted at several public and private universities such as NYU and the University of Texas; these campus

bookstores use software designed by Verba, Inc. to provide students with prices from the campus bookstore and numerous online retailers.

Differences in search behavior across experience levels may be the result of consumers learning about characteristics of the market. However, in order to identify the difference in search behavior that is attributable to differences in consumers' expected value of search, it is necessary to rule out competing hypotheses for the observed difference in search and purchasing behavior. In the college textbook market, this difference could be due to a number of factors that are correlated with experience: who typically pays for the textbook, tastes for quality or retailers, previous online purchasing behavior, the types of courses taken, and beliefs about the value of textbooks among other things. I account for these factors by collecting detailed individual-level data on 6,000 consumers in the University of North Carolina – Chapel Hill (UNC) textbook market using online questionnaires.¹ These data span four semesters and more than 700 independent course-sections. Each observation contains the following information: whether and how the student obtained the assigned textbook, who paid or offered to pay for it, the student's online search behavior and past purchasing behavior, the student's price expectations, and his or her stated preferences regarding bookstore shopping and textbook conditions. In order to evaluate the benefits of search in the online market, I also scrape daily price data from 13 online textbook retailers across two academic calendar years. These data allow me to approximate the empirical distribution of prices across online retailers during the time period considered. I supplement these data with information provided by the university registrar and university bookstore.

In the data I find that inexperienced consumers are significantly less likely to search online retailers than their more experienced counterparts, despite the potential for substantial savings. The price of a new and used textbook from Amazon.com was approximately 30% lower on average than the price of the same textbook from the campus bookstore. This difference in prices was nearly identical for textbooks assigned to first-year students and upper-year students, but first-year

¹Each questionnaire was tailored to a specific course and textbook assignment. Instructors electronically distributed questionnaire links to students in their courses. Student participation was voluntary and encouraged through the opportunity to win cash prizes. This project received approval from the UNC IRB.

students were 18% less likely to search online retailers. Reduced-form results suggest that search behavior is significantly different across experience levels even after controlling for differences in budgets, past experience purchasing online, reported tastes for online and bookstore shopping, and course characteristics.

In order to calculate the value of search across experience levels, I estimate a model of demand with limited information and costly search similar to Hortaçsu and Syverson (2004) and Hong and Shum (2006). In the model, consumers face a decision for one assigned textbook. They have full information about the price and characteristics of the campus bookstore alternatives (new, used, and rental), but face limited information about characteristics of textbooks sold in the online market. In particular, they face uncertainty over prices and their unobservable preference shocks for alternatives in the online market prior to making the search decision (although they know their joint distribution with certainty). Prices and preference shocks are revealed if the consumer pays a known search cost. This search cost is drawn from a distribution that depends on a number of observable factors including previous online shopping experience, experience in the textbook market, and reported confidence in the reliability of the online textbook market.

The majority of the learning literature identifies changes in information by assuming consumers' beliefs follow a Bayesian updating process. In contrast, I identify learning in a reduced-form manner similar to Akerberg (2001). I follow the the majority of other structural models of consumer search in assuming that consumers have rational expectations for prices (i.e., consumers beliefs match the empirical distribution of online prices prior to making the search decision). Consumer learning is captured by reduced form parameters that shift the distribution of search costs. One advantage of this technique is that it avoids the endogeneity of consumers' decisions to acquire signals. The measure of experience used in this paper, semesters enrolled, is arguably exogenous to the decision to search an online retailer for an assigned textbook.² This is contrasted by a measure such as previous online purchases or searches, which are endogenously acquired based on beliefs

²More specifically, as long as consumers are not making decisions about being enrolled in college based on their expectations about the value of searching for a textbook in the online market, then semesters enrolled is exogenous to the search decision for a particular textbook.

about the value of search.³

The search cost and utility parameters are estimated using a modification of a routine proposed by Koulayev (2012). This estimation procedure takes advantage of a reservation utility rule to form choice probabilities used in simulated maximum likelihood estimation. Estimation is facilitated by the observation of both the search and the purchase decision, including the decision to search online retailers and still purchase from the bookstore.⁴

Estimates suggest that experience has a significant effect on the decision to search online retailers, and thus a significant effect on consumers' choice of retailer, even after controlling for consumer, textbook, and course differences. Median estimated search costs are \$13 for inexperienced consumers relative to \$1 for their more experienced counterparts. Counterfactual results suggest that eliminating the gap in search costs across experience levels leads to an average welfare gain of 5% of the price of a new textbook from the campus bookstore (\$8 for a typical textbook assignment).

Overall, these results imply that policies aimed at improving information in the textbook market could significantly benefit consumer welfare by encouraging more search across retailers. Search not only reduces the expected price a consumer will pay, but also induces market participation for marginal consumers who were not willing to pay the pre-search price. In markets where a policy goal is to expand participation, such as alternative energy markets, informational policies that alleviate uncertainty about market-level characteristics are a fiscally responsible way of increasing market participation and total welfare.

These results also have broad implications for firm behavior. Firms may be able to segment

³The decision to acquire signals about the online market could be driven by selection (e.g. individuals who search more are also individuals with lower priors about the distribution of online prices), leading to biased estimates of beliefs and the rate of learning.

⁴A common problem in the search literature is dealing with the issue of endogenous choice set formation. I do not need to use methods to recover latent choice sets because these are observable in the data. The observation of individuals who search the online market and still choose a bookstore alternative also allows for the separate identification of the distribution of search costs versus heterogeneity in tastes for online and bookstore shopping.

markets by experience and take advantage of informational differences to engage in price discrimination. For example, pricing and revision behavior may vary substantially for introductory textbooks which are primarily aimed at inexperienced consumers, than advanced undergraduate textbooks, which are aimed at more experienced consumers). Finally, these results have important implications for antitrust policy in markets with infrequent purchases. When evaluating whether a proposed merger is anticompetitive, antitrust authorities typically consider entry by new firms to be a mitigating factor. My analysis suggests that inexperienced consumers are less likely to search across unfamiliar retailers, representing a barrier to entry for new firms. These information frictions may therefore warrant stricter merger policy in markets where consumers tend to be inexperienced.

The following section provides an overview of research on demand estimation with limited information. Section 3 presents the empirical model. Section 4 presents the data and Section 5 provides parameter estimates and counterfactual results. Section 6 concludes.

2.2 Literature Review

This paper draws from two well-developed literatures in economics: demand estimation incorporating constrained choice sets and learning. The following section describes the relationship between this paper and these literatures and elaborates on how this paper extends these literatures.

2.2.1 Constrained Choice Sets

A critical assumption in a majority of the work done involving demand estimation is that consumers have full information on all product alternatives in a market. Although this assumption is realistic in some markets, this can lead to biased demand estimates in markets where the number of product options is large or if a subset of consumers have not been exposed to various product alternatives. A large amount of work in marketing and economics has focused on how to incorporate limited information over product options into demand estimation. Particular interest has been paid to the determinants of these limited choice sets (e.g., advertising) and to what extent the assumption of full choice sets biases demand estimates. The empirical work done can be separated by two differing approaches: the consideration set approach and the estimation of structural

consumer search models.⁵ Both of these approaches relax the assumption that consumers have full information of all product alternatives and their characteristics, but differ substantially in their methodology. This paper develops and estimates a model that more closely resembles a structural model of consumer search, but the methodology for estimating a more reduced form consideration set model are discussed in the Appendix Section A.1.2.

Incorporating consideration sets into demand estimation began in marketing with Howard and Sheth (1969) and Newell and Simon (1972).⁶ The fundamental premise of the consideration set literature is that when confronted with a purchasing decision, consumers follow a two-step procedure. First, they narrow a full set of alternatives into a smaller, consideration set, and second, they make an optimal choice from this narrower set of alternatives.

In economics, Sovinsky Goeree (2008) uses this methodology to estimate the extent to which the assumption of full information in demand estimation biases estimates of product markups and elasticities in the U.S. personal computer industry. Goeree also examines to what extent advertising affects consumers' choice sets. This paper is similar to Goeree's in that it explores a potential determinant of consumers' information over product options: experience acquired through repeated participation in a market.

A common limitation of the consideration set approach is that it treats individuals' choice sets as being exogenously determined; for example, in Goeree's framework, unobservable preferences for a Dell Inspiron are independent of any unobservables affecting the inclusion of a Dell Inspiron in an individual's choice set. This paper relaxes this limitation by allowing the unobservables in the consideration phase to be correlated with the unobservables in the product choice phase.

⁵This is somewhat of a simplification. A few papers in marketing and economics more closely resemble the work done on estimating structural consumer search models than the majority of empirical work done incorporating consideration sets, but still reference the determination of choice sets as consideration set formation (e.g. Mehta et al. 2003; Pires 2013). In addition to these two approaches, there has been related work by Bruno and Vilcassim (2008) and Conlon and Mortimer (2012) focusing on how to incorporate limited product availability into demand estimation. Although these studies deal with constrained choice sets, they assume consumers have full information over all *available* alternatives.

⁶Van Nierop et al. (2010) provides a survey of the marketing literature on consideration set formation.

Another way of relaxing the assumption of exogenous choice sets is to estimate a more structural model that specifies a decision rule for how consumers choice sets are formed and estimates the parameters that underly the decision rule. This is the approach taken by the literature on the estimation of structural search models.

Despite the wealth of theoretical work done to explain how search costs or frictions affect equilibrium outcomes, no work had been done to structurally estimate consumers' search costs until the early 2000's.⁷ The lack of empirical work on the structural side was primarily due to the difficulty of tracking and quantifying consumers' search behavior. However, the increasing prevalence of internet shopping over the past decade has allowed for detailed data to be collected on consumers' search behavior; with data on search behavior from various internet retailers and "shopbots",⁸ a number of recent papers have been written examining consumer search and price dispersion in online marketplaces (e.g., Hong and Shum, 2006).

The first work to empirically estimate a distribution of consumer search costs was Sorensen (2001). Sorensen uses store level pharmacy data on prices and quantities to estimate parameters of a search model, but because of the limitations of the data, he must make a number of relatively strong assumptions. Other studies have built upon this work to test different models of search (De los Santos et al. 2012b), extend the BLP framework to incorporate endogenous choice sets (Moraga-González et al. 2009), and estimate consumer search costs from observed price dispersion (Hortaçsu and Syverson 2004; Hong and Shum 2006).

These models have two primary advantages over incorporating constrained choice sets through the more reduced form consideration set approach: they explicitly account for the endogeneity of what individuals' choose (or choose not) to gather information about and are able to incorporate uncertainty. For example, Koulayev (2012) uses data on online searching behavior for hotels to

⁷Early empirical work focused on documenting price dispersion in markets with homogeneous goods such as cars and coal (Stigler 1961), gasoline (Marvel 1978), and a variety of other consumer goods (Pratt et al. 1979). Later empirical work focused on using reduced form methods to test results derived from theoretical search models (e.g. Carlson, 1980).

⁸Shopbots are websites devoted to tracking the lowest prices of specific goods from a number of online retailers and then linking consumers to these retailers.

show how the assumption of full information leads to biased results in demand estimation for two reasons: i) because consumers have limited choice sets and ii) because these choice sets are endogenous. In other words, Koulayev shows that, even with knowledge of consumers' choice sets, if the endogenous nature of those choice sets is not accounted for, demand estimation estimates are still biased because consumers choose to gather information (and expand their choice sets) in a non-random manner.

2.2.2 Learning

Previous work on consumer learning has focused on how individuals learn about the specific characteristics of a product (e.g., Erdem and Keane, 1996; Crawford and Shum, 2005). For example, Crawford and Shum (2005), estimate a dynamic model of demand for prescription anti-ulcer medication that incorporates individuals' and doctors' uncertainty about the efficacy of various drugs. They show that initially there is considerable uncertainty over the curative and symptomatic effects of these drugs, but that as individuals repeatedly consume medication, learning takes place relatively quickly.

The majority of studies in the consumer learning literature assume that individuals' beliefs follow some form of Bayesian updating. A notable exception is Akerberg (2001). Akerberg examines the differential effects of advertising on first-time and repeat consumers by incorporating learning into a reduced form demand estimation framework. In his model, consumers learn about their preferences for yogurt indirectly through advertisement (informative effects) and directly through consumption; advertisement also may affect repeat consumers through prestige or image effects. This paper is similar to Akerberg's in the reduced form manner in which learning is accounted for, but I allow for experience to affect individuals' choice sets in addition to the utility they receive from a product.

Two recent papers develop and estimate demand models that incorporate both learning and search: Koulayev (2013) and De los Santos et al. (2012). These papers investigate whether individuals update their beliefs within a search decision. For example, Koulayev uses data on prices and market shares of S&P 500 mutual funds to estimate a model of search that incorporates learning

over the price distribution of funds. This paper differs with these in that it investigates differences in search behavior across purchasing decisions rather than differences in search behavior within purchasing decisions.

2.3 Empirical Model

As students acquire experience in the college textbook market, their observed online search and purchasing behavior both change. Specifically, higher levels of experience are associated with a higher likelihood of searching online retailers, making an online purchase, and also choosing the outside option. To explore the effect of acquired experience on student textbook purchasing behavior, I develop a discrete choice model in which students make an optimal search and purchasing decision, conditional on an instructor's textbook assignment. Experience enters the model by affecting both the decision to search and the decision of which product alternative to choose, conditional on the search decision. This section first outlines individuals' choice sets and how limited information enters the model. Following this, I characterize an individual's optimal search and purchasing decision. I then describe the form of the alternative-specific utility function. The section concludes with a discussion of the simulated maximum likelihood technique used to estimate the parameters of the utility function and search cost distribution.

2.3.1 Full Choice Set

For each course in which an individual is enrolled, individual i is modeled as facing a discrete choice from *at most* 7 alternatives that comprise the full choice set Ω_F .⁹

⁹In practice, a used (or new) book may not be available from the bookstore at the time of purchase. Some assigned books are also not available from online retailers or offered as rentals from the campus bookstore. This exogenous variation in choice sets aids in identification and is accounted for in estimation.

$$\Omega_F = \left\{ \begin{array}{l} 0 \text{ Do not purchase a textbook (outside option)} \\ 1 \text{ Purchase a new textbook from a bookstore} \\ 2 \text{ Purchase a used textbook from a bookstore} \\ 3 \text{ Rent a textbook from a bookstore} \\ 4 \text{ Purchase a new textbook from an online retailer} \\ 5 \text{ Purchase a used textbook from an online retailer} \\ 6 \text{ Rent a textbook from an online retailer} \end{array} \right.$$

Alternatives are partitioned by type (new, used or rental¹⁰) and by retailer location (physical bookstore or online retailer). The choice to not purchase a textbook (the outside option) also includes the choice to use a reserve copy at the library, borrow a copy from another student, or purchase or rent a previous edition of the assigned textbook.

2.3.2 Limited Choice Set

A key feature of the college textbook market is that consumers are initially steered towards a common set of purchasing options: those offered at the university bookstore. Because of this, I model consumers as having full information about the bookstore options prior to the purchasing decision.¹¹ With this information in hand, consumers then make a decision to obtain information about alternatives available from online retailers, Ω_O . If individual i decides not to search online retailers for information about the assigned textbook, her choice set is constrained to Ω_L , otherwise

¹⁰Students are able to rent books for the duration of a semester at a lower price than the price to purchase a used book. If the book is not returned, the student is generally forced to pay the difference between the new and rental price. Also, the decision to rent is not separated into the decision to rent a new or used option primarily because the majority of online retailers do not make the quality of the rental option explicit.

¹¹It is assumed that each student's choice set contains the bookstore options for a number of reasons: students are directed to the campus bookstore website in order to find the assigned book(s) for their courses, where prices are listed adjacent to the title and ISBN of the assigned book(s), information on the campus bookstore is presented at orientation and students receive periodical emails from the campus bookstore.

she chooses the alternative with the maximal utility in the full choice set, Ω_F :¹²

$$\Omega_F \left\{ \begin{array}{l} \Omega_L \left\{ \begin{array}{l} 0 \text{ Do not purchase a textbook (outside option)} \\ 1 \text{ Purchase a new textbook from a bookstore} \\ 2 \text{ Purchase a used textbook from a bookstore} \\ 3 \text{ Rent a textbook from a bookstore} \end{array} \right. \\ \Omega_O \left\{ \begin{array}{l} 4 \text{ Purchase a new textbook from an online retailer} \\ 5 \text{ Purchase a used textbook from an online retailer} \\ 6 \text{ Rent a textbook from an online retailer} \end{array} \right. \end{array} \right.$$

2.3.3 Search Decision

Consumers choose to search online retailers if the expected benefit of search outweighs an individual specific search cost. Let u_{ij} denote the alternative-specific utility individual i receives from choice j . Further, let $U_{iL}^* = \max\{u_{ij} : j \in \Omega_L\}$, $U_{iO}^* = \max\{u_{ij} : j \in \Omega_O\}$, and $U_{iF}^* = \max\{u_{ij} : j \in \Omega_F\}$ denote individual i 's utility from the alternative yielding the highest utility in the limited, online, and full choice sets respectively. If individuals are allowed to freely revisit the alternatives in the limited choice set (i.e. search with recall), the ex-post benefit of search to individual i is then,

$$U_{iF}^* - U_{iL}^* \tag{2.1}$$

¹²The assumption that some individuals do not have the online options in their choice set does not necessarily imply that these individuals are unaware that textbooks can be purchased through online retailers. Instead, it implies that individuals who do not price the assigned textbook at an online retailer do not have the information necessary to purchase an online alternative and that these alternatives are not in their choice set.

or equivalently,

$$U_{iO}^* - U_{iL}^* \quad \text{if } U_{iO}^* > U_{iL}^* \quad (2.2)$$

$$0 \quad \text{if } U_{iO}^* < U_{iL}^* \quad (2.3)$$

Equation (2.1) states that the ex-post benefit of search is simply the maximal utility from all alternatives less the maximal utility from only the bookstore alternatives and the outside option (i.e. the limited choice set). Before considering the ex-ante benefit of search, it is helpful to rewrite Equation (2.1) into the two cases given by Equations (2.2) and (2.3). If the maximal utility from the online alternatives is greater than the maximal utility from the bookstore alternatives and the outside option, then the ex-post benefit of search is simply the difference between the two. In the other case where the maximal utility from the online alternatives is less than the alternatives in the limited choice set, then the ex-post benefit of search is zero.

The maximal utility of the limited choice set, U_{iL}^* , is assumed to be known to the individual before search, but the individual faces uncertainty about characteristics of the online alternatives that search reveals. In particular, individuals face uncertainty about the prices and additional unobservable preference shocks of the online alternatives that is realized only after search.

The expected benefit of search is,

$$\int_{U_{iL}^*}^{\infty} (U_{iO}^*(p_o, \epsilon_o) - U_{iL}^*) dG(U_{iO}^*) \quad (2.4)$$

G is the cumulative distribution function of individual i 's beliefs about the maximal utility from the online alternatives. U_{iO}^* is explicitly written as a function of p_o and ϵ_o – vectors of the prices of the online alternatives and unobservable (to the individual and econometrician) preference shocks – the only random components in the above equation. The lower bound on the expected benefit from search is U_{iL}^* because the decision to search does not exclude options from an individual's

choice set.¹³

Consumers' search costs are assumed to be known at the time of search, which yields the following search decision rule, where an individual searches if the expected benefit outweighs an individual specific search cost, c_i :

$$\int_{U_{iL}^*}^{\infty} (U_{iO}^* - U_{iL}^*) dG(U_{iO}^*) > c_i \quad (2.6)$$

Following the search decision, an individual chooses the alternative yielding the highest utility, conditional on that individual's choice set.

In order to estimate the model, assumptions need to be made on an individual's search costs, c_i and beliefs about the distribution of U_{iO}^* . Search costs, c_i , are parametrized as being drawn from a population lognormal distribution:

$$c_i \sim \ln \mathcal{N}(\gamma' W_i, \theta^2) \quad (2.7)$$

Explicitly, W_i contains,

- Semesters and semesters squared terms.
- An indicator equal to one if individual i has an older sibling who attended college previously.
- An indicator equal to one if individual i has made *any* online purchase before.
- Stated preferences for confidence in online shopping.

Each covariate in W_i except for the semester indicators are assumed to be exogenous search cost shifters and are excluded from the specification of u_{ij} . For example, an individual who has previously made an online purchase may be more familiar with online shopping or may have a Paypal.com or Amazon.com account, reducing the search costs necessary to purchasing a textbook

¹³Equation (2.4) can be rewritten as

$$\int_{p_o} \int_{\epsilon_o} (U_{iO}^*(p_o, \epsilon_o) - U_{iL}^*) 1[U_{iO}^* > U_{iL}^*] dH(\epsilon_o) dF(p_o) \quad (2.5)$$

where p_o is the vector of prices of the online alternatives and ϵ_o is a vector of online alternative-specific unobservables.

online. However, this is assumed to not affect how much an individual values a textbook from the internet versus a bookstore conditional on the search decision or the valuation of a textbook at the extensive margin. In other words, having previously made an online purchase is assumed to only affect the process of an individual searching.

Recall that individuals face uncertainty over the prices of the online alternatives (new, used, and rental) and also other unobservable factors realized after search. Unobservable demand shocks are assumed to be distributed Type 1 Extreme Value¹⁴ and observed empirical price distributions are used to reflect consumers' beliefs about the price distribution of the online alternatives (i.e. rational expectations).¹⁵

2.3.4 Alternative-Specific Utility

Student i receives utility from choice j ,

$$u_{ij} = \beta_j' X_{ij} + \alpha_i p_{ij} + \epsilon_{ij} \quad (2.8)$$

The utility from choice 0 (not purchasing a textbook) is normalized to $u_{i0} = \epsilon_{i0}$. For certain covariates in X_{ij} , the value of β only varies by the outside option.¹⁶

In order to control for differences across the characteristics specific to each alternative, X_{ij} contains the following:

- Indicators for new, used, or rental alternatives.

¹⁴Further analysis could relax this assumption by using a Nested Logit or Generalized Extreme Value Distribution similar to that in Bresnahan et al. (1997)

¹⁵If consumers have upward biased beliefs about the distribution of prices from online textbook retailers, this will bias estimated search costs upward and price sensitivities towards zero (upwards biased price beliefs, high search costs, and low price sensitivities could all lead to lower levels of search). Matsumoto and Spence (2013) have found evidence that inexperienced consumers' price expectations are significantly higher than their more experienced counterparts using a novel dataset on consumers' subjective beliefs about the distribution of prices of textbooks from online retailers. Thus, a limitation of this paper is that any differences in beliefs across experience levels will cause search costs to be biased upwards for inexperienced consumers. Although this is a limitation, it does not diminish the counterfactuals of this paper. Instead, it simply does not allow differences in search costs across experience level to be differentiated from differences in price beliefs.

¹⁶In other words, for certain covariates in X_{ij} , $\beta_j = \beta_k \quad \forall j, k \in \{1, 2, \dots, 6\}$. These restrictions are discussed in the results section.

- Indicators for bookstore and online alternatives.
- A vector of stated preferences for quality, Z_{1i} , interacted with the new indicator.¹⁷
- A vector of stated preferences for bookstore shopping, Z_{2i} , interacted with the bookstore indicator.

X_{ij} also contains the following individual, textbook, and course characteristics:

Individual Characteristics:

- Semesters and semesters squared terms.¹⁸
- Indicators for on-campus students, in-state students, and if the course is in the student's major field.

Textbook Characteristics:

- Indicators for guaranteed buyback textbooks,¹⁹ required books (as opposed to recommended), books with bundled materials, hardcover books, and UNC-custom editions or coursepacks.

Course Characteristics:

- Total enrollment in the course.
- Indicators for upper level courses, courses in the spring, and if the instructor informed the individual of purchasing options before the semester began.

The covariates above are included to account for individual, course, and textbook characteristics that could affect purchasing and search behavior and may be correlated with market experience. For example, upper-level students may be observed to search online retailers more because the informal secondary market (e.g., student to student) is different for a typical upper-level course than a typical introductory course. This necessitates the inclusion of controls for upper-level and large enrollment courses.

¹⁷A detailed description of how these stated preferences were collected and measured can be found in Section 2.4.1.

¹⁸In future versions of this paper, I plan to experiment with various functional forms of semesters enrolled.

¹⁹The campus bookstore offers consumers a guarantee that they will buy their textbook at a specified price for a small number of large enrollment courses.

Experience enters the utility function through the semesters and semesters squared term in X_{ij} .²⁰ These terms reflect changes in how students value textbooks as they accrue experience.²¹ For example, a first-semester student may have different beliefs before the semester begins about how much a textbook will be used during a semester. As the student progresses through college, she may refine that expectation, leading her to change her valuation of an assigned textbook. This change could also reflect differences in the value of the outside option as a student progresses through school. Consumers may acquire information about reserve copies at the library or the ability to purchase previous editions as they gain experience. If these options are preferable to purchasing/renting a current edition of the assigned textbook, this will be reflected in a change in the value of a textbook at the extensive margin.²²

An additional covariate in X_{ij} measures the utility individuals receive from the continuation value of the textbook at the end of the semester.²³ This continuation value is given by the term $\max\{r_{ij}, \bar{r}_{ij}\}$. The expected resale value of choice j to student i at the end of the semester is given by r_{ij} (where the expectation is taken at the beginning of the semester), and \bar{r}_{ij} is student i 's expected reservation resale price (the lowest price a student expects she would be willing to sell her textbook for at the end of the semester).²⁴ If $r_{ij} > \bar{r}_{ij}$ the expected resale price is greater than the expected reservation price and the individual intends to sell the book at the end of the

²⁰The parameters associated with semesters enrolled in the utility function are separately identified from the semesters enrolled terms in the search cost distribution according to the following logic. Unconditional differences in the decision to search online across semesters enrolled identify the terms in the search cost distribution. The terms in the utility function are identified through variation in the decision to purchase a textbook at the extensive margin across semesters *conditional* on an individual searching. For this reason, the observation of both the purchasing decision and the search decision is critical for identification.

²¹Semesters are measured as the number of semesters a student has attended college *before* the semester in which the choice is made (i.e. the minimum value of $sem_i = 0$).

²²It should be noted that this reduced form measure of learning is not the focus of the paper. Experience enters the utility function due to the observed trend that upper-level students search online retailers more *and* choose the outside option more frequently than first-year students. This can only be rationalized if the outside option is becoming more attractive with experience.

²³This covariate is excluded from the rental alternative as well as the outside option.

²⁴Both r_{ij} and \bar{r}_{ij} are elicited through questions in the online questionnaire (i.e., both are data rather than latent variables).

semester. If $\bar{r}_{ij} > r_{ij}$, the student expects that the value of the textbook to be greater than the resale price and plans to keep the textbook. Although the *max* operator does not fully characterize a individual's optimization problem with respect to the continuation value of the textbook at the end of the semester – it only depends on the first moments of these random variables – it provides a intuitively reasonable and tractable approximation.

Other terms in the utility function include the log price of choice j , given by p_{ij} . The price sensitivity parameter varies by four deterministic cases and are given by α_i . Explicitly,

$$\alpha_i = \alpha_0 + \alpha_1 S_{1i} + \alpha_2 S_{2i} + \alpha_3 S_{3i} \quad (2.9)$$

The model allows for four deterministic price sensitivities: i) the student purchases the textbook with his or her own money, ii) the student receives a fixed payment from another person in order to purchase the textbook, iii) the student receives a varying payment from another person, and iv) the student pays for the textbook with funds from a scholarship. An example of the second case is another person or scholarship providing an individual with \$X for textbooks allowing the individual to keep any potentially remaining money. The third case corresponds to the case of another person completely covering the cost of textbooks regardless of the price. In the specification given by Equation (2.9), S_{1i} is an indicator equal to 1 if the individual received a varying payment; S_{2i} is an indicator equal to 1 if the individual received a varying payment; S_{3i} is an indicator equal to 1 if the individual received scholarship money; the omitted category is paying out-of-pocket.

2.3.5 Reservation Utility Rule

In order to write out the likelihoods of the observed product and search decisions, it is first helpful to characterize the decision rule in Equation (2.6) in terms of reservation utility in a manner proposed by Koulayev (2012). An individual will choose to search if the alternatives in the limited choice set provide utility below a threshold value, \bar{u}_i , at which the individual is indifferent between the decision to search and not to search. Explicitly, an individual will search if,

$$U_{iL}^* < \bar{u}_i \quad (2.10)$$

where the reservation utility, \bar{u}_i , is the utility level where the expected benefit from search is equivalent to the cost of search:²⁵

$$\int_{\bar{u}_i}^{\infty} (U_{iO}^* - \bar{u}_i) dG(U_{iO}^*) = c_i \quad (2.11)$$

The LHS of Equation (2.11) can be approximated using simulation methods for any value of \bar{u}_i . In practice this is done using the following algorithm,

1. Begin with a guess of \bar{u}_i .
2. Take NS_1 draws from a standard uniform distribution²⁶: $\mu \sim U(0, 1)$. Each individual draw will be denoted with a superscript s (e.g., μ^s is a scalar draw from standard uniform).
3. Take NS_2 draws from the empirical distribution of prices for each online alternative (new, used, rental). Each draw from the price distributions is denoted as the vector p_O^s .
4. Calculate the inverse CDF of U_{iO}^* for each μ^s .

$$G^{-1}(\mu^s) = \sum_{k=1}^{NS_2} \left[-\ln \left(\frac{-\ln(\mu^s)}{\sum_{j \in \Omega_O} \exp(v_{ij}(p_{ij}^k))} \right) \right] \quad (2.12)$$

where $v_{ij} = u_{ij} - \epsilon_{ij}$. For each s , this is essentially taking a draw from U_{iO}^* . Note that v_{ij} is a function of the online prices, p_O , for the online alternatives.

5. If the draw from U_{iO}^* is greater than \bar{u}_i , then assign y^s as the expected benefit of search for one simulation draw, equal to the draw for U_{iO}^* minus the value of \bar{u}_i , otherwise $y^s = 0$:

$$y^s = \begin{cases} G^{-1}(\mu^s) - \bar{u}_i & \text{if } \mu^s > G(\bar{u}_i) \\ 0 & \text{if } \mu^s \leq G(\bar{u}_i) \end{cases}$$

²⁵The function G is used for notational simplicity. In practice, integrating over individuals' beliefs about the benefit of search is a six dimensional integral over the three online epsilon shocks and three different online prices (new, used, and rental). The type-1 extreme value assumption allows for the three dimensional integral over the epsilons to be reduced to a one dimensional integral.

²⁶Draws from standard uniform distributions used for simulation are taken from Halton sequences. Each sequence is formed using a different prime number and the twenty initial draws from each sequence are discarded (to avoid correlation across sequences).

6. The simulated value of $\int_{\bar{u}_i}^{\infty} (U_{iO}^* - \bar{u}_i) dG(U_{iO}^*)$ is given by,

$$\frac{1}{NS_1} \sum_{s=1}^{NS_1} y^s \quad (2.13)$$

Using this technique \bar{u}_i can be solved for by finding the root of $\int_{\bar{u}_i}^{\infty} (U_{iO}^* - \bar{u}_i) dG(U_{iO}^*) - c_i$ for any value of c_i . Choice probabilities are formed based on these individual specific reservation utilities.

2.3.6 Likelihood Function

The individual likelihoods can be separated into the three following cases:

- The likelihood of an individual not searching and choosing the outside option or a bookstore option, $j \in \{0, 1, 2, 3\}$, is

$$Pr_{ij}^{NS} = Pr(u_{ij} > u_{ik}, \forall k \neq j \in \Omega_L) \cdot Pr(U_{iL}^* > \bar{u}_i) \quad (2.14)$$

- The likelihood of an individual searching and choosing the outside option or a bookstore option, $j \in \{0, 1, 2, 3\}$ is

$$Pr_{ij}^S = Pr(u_{ij} > u_{ik}, \forall k \neq j \in \Omega_L) \cdot Pr(U_{iL}^* < \bar{u}_i) \cdot Pr(U_{iO}^* < U_{iL}^* | U_{iL}^* < \bar{u}_i) \quad (2.15)$$

- The likelihood of an individual searching and choosing an online or rental option, $j \in \{4, 5, 6\}$ is

$$Pr_{ij}^S = Pr(u_{ij} > u_{ik}, \forall k \neq j \in \Omega_O) \cdot Pr(U_{iL}^* < \bar{u}_i) \cdot Pr(U_{iO}^* > U_{iL}^* | U_{iL}^* < \bar{u}_i) \quad (2.16)$$

where Ω_O is the choice set composed of only the online and rental options.

Equation (2.14) states that the probability of an individual not searching and choosing either the outside option or a bookstore option is the choice probability conditional on choosing from

the limited choice set times the probability that the maximal utility from the limited choice set is greater than the individual's reservation utility.

Equations (2.15) and (2.16) include an additional term because the product choice conditional on choosing to search provides information about both U_{iL}^* and U_{iO}^* . The second term in equation (2.15) indicates that individual searched because the maximal element in U_{iL}^* was less than \bar{u}_i . Because the individual chose an element from Ω_L , it must be the case that $U_{iL}^* > U_{iO}^*$ (the third term in Equation (2.15)). The first term is simply the probability the individual chose alternative j in the limited choice set.

The first terms of Equations (2.14) - (2.16) have a closed form expression. The second terms of these equations have a closed form representation conditional on a value of \bar{u}_i . The third term of Equations (2.15) and (2.16) must be approximated using simulation methods.

Using these choice probabilities, the parameters of the alternative specific utility function and search cost distribution, collectively Θ , can be estimated using simulated maximum likelihood estimation. The unconditional individual likelihood function is given by,²⁷

$$\mathbb{L}_i(\Theta) = \int_c \int_{p_o} \left[\prod_j (P_{ij}^S)^{d_{Sij}} (P_{ij}^{NS})^{d_{NSij}} \right] dH(c) dK(p_o) \quad (2.17)$$

for $j \in \{0, \dots, 6\}$. The distribution of search costs is given by the function H . The multivariate distribution of online prices is given by the function K . Individuals are assumed to know the empirical distribution of online prices (i.e., rational expectations). The indicator function, d_{Sij} (d_{NSij}), is equal to one if the individual chooses to search (not search) and chooses alternative j . This individual likelihood can be approximated using simulation techniques. Each observation is treated as independent; the full likelihood function used in estimation is given by,

$$\mathbb{L}(\Theta) = \prod_i \mathbb{L}_i(\Theta) \quad (2.18)$$

²⁷The probability of not searching and choosing an online alternative, P_{ij}^{NS} for $j \in \{4, 5, 6\}$, is zero.

The Appendix Section A.1.1 provides more detailed information on the form of the above choice probabilities, how the probabilities without closed forms are approximated, and the form of the likelihood function.

2.4 Data

2.4.1 Online Questionnaire Data

Micro-data were collected through online questionnaires sent from instructors to their students.²⁸ At the beginning of the fall and spring semesters for the 2011-2012 and 2012-2013 academic calendar years, instructors were sent an email containing survey directions and a link to an online survey. This email was then forwarded by the instructor to the students in his or her class. Each survey was individualized to the specific course; students were asked if and how they obtained the assigned textbook for the course, price information, which retailers they considered, details of their online search behavior, their price expectations of retailers not considered, past purchasing behavior, characteristics of the textbook they (may have) purchased, and questions gauging preferences for bookstore shopping and quality among other items.²⁹

For the 2011-2012 and 2012-2013 academic calendar year, instructors of more than 700 course-sections forwarded an email containing an online survey link to their students.³⁰ This resulted in 8,279 students responding. Of the students who began the survey, 7,191 completed the survey: an 87% completion rate.³¹

The summary statistics that follow exclude non-undergraduate students (248 excluded), respondents whose textbooks were available for free through the UNC student stores because of scholarships offered through the university (315 excluded), or did not have information for other

²⁸UNC IRB study #11-1177. The online surveys are conducted through the Odum Institute.

²⁹The Appendix Section A.2.2 contains a detailed description of the online questionnaire.

³⁰These surveys were distributed to instructors following the last day students were able to return textbooks to the campus bookstore for a full refund; this is roughly two weeks after the start of the semester. In an effort to bolster student response rates, respondents were entered in a drawing to win cash prizes.

³¹1,892 students responded in the fall of 2011, 2,493 students responded in the spring of 2012, 2,189 students responded in the fall of 2012, and 1,705 students responded in the spring of 2013.

key variables (595 excluded).³² These individuals are also excluded in estimation, leaving a sample of 6,033 students.

Table 2.1 summarizes the distribution of respondents' number of semesters enrolled. The number given in Table 2.1 includes the current semester of the respondent, as well as summer sessions and semesters enrolled at any other university.

Table 2.2 summarizes the demographics of the respondents. This includes the percentage of students who were classified by the university as in-state students, live on campus, male, and were the first child in their immediate family to attend college. This table also includes the age of respondents and the percentage of students who reported that the course in which they were surveyed was in their major field. The total number of responses varies because response to these questions was optional.

Table 2.3 summarizes respondents' textbook purchasing decisions; students are separated into two categories: freshmen (defined as enrolled in two or fewer semesters) and upperclassmen. For freshmen, 84% of students decided to purchase the assigned textbook, 5.4% chose to rent the assigned textbook, 5.8% borrowed the textbook from another student, and a small percentage chose to use a copy from the library or to not use the textbook entirely. Upperclassmen tended to purchase the book less frequently, rent more frequently, and chose the other options at roughly the same rate.

Table 2.4 summarizes students' choice of retailers if they purchased the assigned textbook. Observations for courses that assigned textbooks which were not available at online retailers (UNC Coursepacks or Custom Editions) are excluded from this table. Conditional on being able to purchase the assigned textbook through an online retailer, the data show that upperclassmen were more likely to choose to purchase the assigned textbook through an online retailer than freshmen. Perhaps surprisingly, freshmen were more likely to purchase the assigned textbook from another student than upperclassmen. These purchases occurred almost exclusively in the spring semester and appear to be a result of the type of course the freshmen were taking (the supply of textbooks

³²This last number is higher than it will be in the final version of the paper. Individuals who choose an "other" option for key questions are dropped in this version, but because they were prompted to write in responses, imputations can be made for the choices they made.

for introductory courses through other students is larger than for more specialized, core courses).

Table 2.5 summarizes students choice of online retailer, conditional on purchasing or renting online. The majority of students purchased their textbook through Amazon.com, accounting for more than seven times the purchases of the next leading online retailer, Half.com. Roughly half of all rentals were done through Chegg.com (although this still represents a small proportion of the total number of online purchases and rentals). HTML source codes were collected for eight additional online retailers and two publishing websites, which comprised an additional 8.6% of sales and rentals. Price data were not gathered for 6.8% of the observed sales and rentals (e.g. alibris.com, bookbyte.com).

Table 2.6 contains information on whether students knew about the possibility of renting textbooks at the beginning of the semester and whether they priced the textbook at another physical bookstore, or an online retailer. Evidence that students learn about the availability of product options is seen in students' knowledge of rental options at the beginning of the semester; the vast majority of upperclassmen surveyed were knowledgeable of rental options when making their textbook decision, while a sizable fraction of freshmen were not.³³

This table also includes the proportion of students who had ever purchased *anything* online or ever made an online sale. Nearly all students, freshmen and upperclassmen, had made an online purchase before, but less than half had ever made an online sale before.

Table 2.7 summarizes students' stated preferences for aspects important to the decision to buy a new or used textbook and to deciding whether to buy from a physical bookstore or an online retailer.³⁴

These stated preferences include responses to the following questions important to the bookstore/online retailer decision:

³³In the model, all individuals have full information about the availability of the rental options.

³⁴Respondents used a sliding scale to report their preferences, with response values ranging from 0 to 10. For example, for the question "How important to you is buying a new rather than a used textbook?", the boundary at zero is labeled, "I am completely indifferent between buying a new and a used textbook." The boundary at a response of ten is labeled, "I will only buy a new textbook." In order to alleviate measurement error decimal responses to the nearest tenth were allowed. See the Appendix Section A.2.2 for a screenshot of these questions.

- How important is seeing the condition of a textbook before purchasing it?
- How important is receiving a textbook on the day you purchase it?
- How confident are you that you will be satisfied with an online purchase?

Questions gauging the importance of buying a new rather than a used textbook include the following:

- How important is purchasing a new rather than a used textbook?
- How important is purchasing a textbook without highlighting, dog-ears, or notes already in it?

Students reported placing low importance on purchasing a new rather than a used textbook. Slightly more importance was placed on purchasing a textbook without notes or highlighting in it, seeing the textbook before purchasing it, and receiving the textbook on the day of purchase. The majority of students reported being confident in being satisfied with an online purchase.

Table 2.8 reports the correlation between stated preferences.³⁵ Correlations between stated preferences are of the expected signs. Preferences for purchasing a new rather than a used textbook are positively correlated with preferences for purchasing a textbook with no notes or highlighting. These preferences are also correlated with seeing the book before purchasing it, but to a lesser degree. Preferences important for the bookstore/online retailer decision, the importance of receiving the textbook on the same day and seeing the textbook before purchasing it are also highly correlated.

As expected, confidence in online shopping is negatively correlated with preferences for seeing the textbook and receiving the textbook on the same day. Interestingly, freshmen seem to have stronger preferences for aspects of textbook purchasing that would lead to more new rather than used purchases and more bookstore than online purchases; this lends support to the need to include these measures when examining the effects of experience on purchasing behavior.

³⁵P-values are indicated in parentheses.

Table 2.9 summarizes who paid for the assigned textbook. Less than half of all students paid for the assigned textbook out of pocket, with slightly more upperclassmen paying for the textbook than freshmen. Of the students who received money from another person to pay for the textbook, the majority received a payment that varied with the cost of the textbook (i.e. parents paid for the entire cost of textbooks, regardless of cost). A much smaller proportion received a fixed payment from another person to cover the cost of the textbook.

2.5 Results

This section first presents parameter estimates for the alternative specific utility function and search cost distribution. This is followed by a discussion of estimated elasticities. I conclude this section with a discussion of model fit and counterfactual results.

2.5.1 Parameter Estimates

Recall that search costs are distributed log-normal according to the following,

$$c_i \sim \ln \mathcal{N}(\gamma' W_i, \theta^2) \quad (2.19)$$

Table 2.10 presents estimates for the parameters that determine the mean and variance of the search cost distribution. Consumers with more tenure in the market have significantly lower search costs, although the marginal reduction in search costs from an additional semester is decreasing. The parameters given below imply that the median search cost for a first year student is \$13.33 and \$1.34 for an upper-level student.

Consumers who have made a previous online purchase or have higher levels of stated confidence in online shopping also have significantly lower search costs. Consumers who have not had older siblings in college before them have search costs that are not statistically different than those with with siblings in college. This suggests that there are not significant spillover effects from having a close relationship with a more experienced market participant, which is also suggestive that personal experience is driving the differences in search costs rather than peer effects.³⁶

³⁶The variance of the search cost distribution is very high. Although median search costs are reasonable, mean

Table 2.11 presents the parameter estimates for the alternative-specific utility function. Consumers value textbooks less as they progress through school, for courses in their major field, and courses in the spring. This result is surprising for students' purchasing textbooks within their major fields. One possible explanation is that the negative coefficient reflects multicollinearity between students taking courses within their major field and semesters enrolled (upper year students are more likely to be taking courses in their major field). Stated preferences associated with valuing quality and newness are associated with greater valuation of new textbooks.

Consumers paying for their textbook out of pocket are more price sensitive than consumers receiving fixed or varying payments. Consumers that report paying for their textbook with money from a scholarship are the most price sensitive group. This potentially reflects differences in these students budget constraints. Unfortunately, I do not know whether the scholarship funds are for merit- or needs-based reasons.

2.5.2 Elasticities

Table 2.12 presents simulated price elasticities. Demand for the bookstore alternatives is inelastic. This implies that bookstore profits would unambiguously increase from an increase in prices. This counterintuitive result is likely due to features of the textbook market. The first feature is that the campus bookstore is not necessarily profit maximizing. The campus bookstore is designated as a non-profit organization and a portion of its revenues are used to support non-athletic scholarship and fellowship programs of the university.³⁷

Demand is also inelastic for the online alternatives. This implies that if all online retailers and marketplace sellers raised their prices, overall profits would be increased. However, the price elasticity of demand for an individual seller on an online marketplace is likely much more elastic (relative to the entire market).

search costs are relatively large. Additional analysis would include checking the sensitivity of the log-normal distribution by using other distributions with positive support (e.g., Gamma).

³⁷In addition to this, there is already public sentiment against the rising costs of textbooks at the campus bookstore. In 2006, a UNC subcommittee was formed to review the rising costs of textbooks. It's possible that the bookstore does not charge the static profit maximizing price to reduce future objections about rising textbook prices (and potential negative publicity).

Cross price elasticities are sensible. Consumers that purchased from the campus bookstore are more likely to substitute to another bookstore purchase if the price of their original choice increases. Similarly, cross price elasticities within the online alternatives is larger than the cross price elasticities between the online alternatives and bookstore alternatives.

2.5.3 Counterfactuals and Model Fit

The first column of Table 2.13 reports the simulated choice probabilities for the baseline model. The second column reports the simulated choice probabilities for a counterfactual where all consumers' search costs are drawn from the same distribution as eighth semester consumers (while holding their other observable characteristics constant). In other words, all consumers now have the same beliefs about the value of searching in the online market, conditional on other observables being the same. In this counterfactual, inexperienced consumers decrease the frequency with which they choose the outside option by sixteen percent, reduce bookstore purchases by approximately sixteen to eighteen percent, and increase online purchases/rentals by fifteen to sixteen percent.

The goal of this counterfactual is to calculate the welfare gain to inexperienced consumers from the reduction in search costs. Inexperienced consumers can be broken into three groups: i) those who did not search prior to or following the reduction in search costs (24%), ii) those who searched prior to and following the reduction in search costs (63%), and iii) those who did not search prior to the reduction and did search following the reduction (13%). For groups i) and ii), there are no welfare gains from the reduction in search costs. In the third group, there are some individuals who are induced to search, but still choose their original alternative (prior to the counterfactual). There are also no welfare gains to this group. However, there are some consumers in the third group that are induced to search and switch from a bookstore alternative or outside option to an online alternative (66% of the third group). For this group, there is a positive welfare gain. Averaging across all first-year students (those with zero and positive welfare gains) results in an average welfare gain of \$8.49 for inexperienced consumers. Conditional on having a positive welfare gain, the gain is \$102.09 on average.

2.6 Conclusion

This paper finds significant differences in consumers' search and purchasing behavior across levels of market experience. After controlling for consumer, course, and textbook characteristics that are correlated with experience, consumers' decision making is still markedly different across experience levels due to changes in consumers' expected value of search. These results imply that policies that provide information to consumers about characteristics of the online textbook market would yield significant benefits to inexperienced consumers. Moreover, these results carry additional weight in markets where public policy is concerned with increasing access (participation) in the market. I find that information leads to more search which induces market participation for marginal consumers who were not willing to pay the pre-search price. In markets with a significant fraction of inexperienced consumers, informational policies that alleviate uncertainty about market-level characteristics may be a fiscally responsible way of increasing market participation.

This paper takes advantage of an exogenous measure of experience in the college textbook market (semesters enrolled), but there are tradeoffs associated with using this measure. I am able to capture average changes across consumers as gain experience in the market, but am not able to identify the precise signals that consumers are receiving. There are a number of potential signals that consumers receive in this market: past searches or interactions with other market participants (peer effects) are two possible sources. This paper is further limited in its identification of the characteristics of the market that consumers are learning about. In a separate paper with Brett Matsumoto, we find that inexperienced consumers have biased expectations about the price of textbooks in the online market, but that these expectations tend to converge to the empirical mean with experience (Matsumoto and Spence 2013).

A natural followup question to this paper is whether firms are able to extract surplus from inexperienced consumers in the form of price discrimination or some other mechanism (e.g., revisionary behavior). Understanding firm responses to heterogeneity in consumer experience is important for policies related to both improving consumer outcomes and developing anti-trust policy. In markets such as durable goods markets, a large proportion of consumers will have made

relatively few purchases. If inexperienced consumers are less likely to search unfamiliar or new retailers, than this represents a barrier to entry for new firms. These information frictions may therefore warrant stricter merger policy in markets where consumers tend to be inexperienced.

Table 2.1: Semesters Enrolled

| Semesters | Frequency | Percent | Cumulative |
|-----------|-----------|---------|------------|
| 1 | 821 | 13.6 | 13.6 |
| 2 | 789 | 13.1 | 26.7 |
| 3 | 614 | 10.2 | 36.9 |
| 4 | 782 | 13.0 | 49.8 |
| 5 | 681 | 11.3 | 61.1 |
| 6 | 705 | 11.7 | 72.8 |
| 7 | 617 | 10.2 | 83.0 |
| 8 | 552 | 09.1 | 92.2 |
| 9+ | 472 | 07.8 | 100 |
| Total | 6,033 | - | - |

Table 2.2: Student Demographics

| | N | Mean | Std. Dev. | Min | Max |
|-----------------------|------|------|-----------|-----|-----|
| In-State | 6028 | 0.79 | - | 0 | 1 |
| On-Campus | 6026 | 0.60 | - | 0 | 1 |
| Male | 6022 | 0.38 | - | 0 | 1 |
| First in College | 6033 | 0.54 | - | 0 | 1 |
| Course in Major Field | 6033 | 0.59 | - | 0 | 1 |
| Age | 5978 | 19.9 | 2.1 | 16 | 61 |

Table 2.3: Students' Beginning of Semester Textbook Decisions

| Option | Freshmen | | Upperclassmen | |
|------------------------------|-----------|---------|---------------|---------|
| | Frequency | Percent | Frequency | Percent |
| Purchased the Textbook | 1,353 | 84.0 | 3,330 | 75.3 |
| Rented the Textbook | 87 | 05.4 | 450 | 10.2 |
| Borrowed the Textbook | 94 | 05.8 | 262 | 05.9 |
| Used a Copy from the Library | 8 | 00.5 | 101 | 02.3 |
| Did not use the Textbook | 68 | 04.2 | 280 | 06.3 |

Table 2.4: Students' Retailer Choices

| Retailer | Freshmen | | Upperclassmen | |
|---------------------------|-----------|---------|---------------|---------|
| | Frequency | Percent | Frequency | Percent |
| UNC Student Stores | 485 | 53.1 | 1,172 | 39.0 |
| Another College Bookstore | 4 | 00.4 | 61 | 02.0 |
| Online Retailer | 359 | 39.3 | 1,641 | 54.7 |
| Another Student | 65 | 07.1 | 127 | 04.2 |

Table 2.5: Students' Online Retailer Choices - Purchased or Rented

| Choice of Online Retailer | Frequency | Percent |
|----------------------------|-----------|---------|
| Amazon | 1,613 | 66.9 |
| Half (Ebay) | 226 | 09.4 |
| Chegg | 203 | 08.4 |
| Other (Data Collected) | 208 | 08.6 |
| Other (Data Not Collected) | 163 | 06.8 |

Table 2.6: Students' Purchasing Behavior

| | Freshmen Percent | Upperclassmen Percent |
|------------------------------|---------------------|--------------------------|
| Knowledge of Rental Options | 74.9 | 93.1 |
| Priced an Online Retailer | 61.0 | 74.0 |
| Ever Made an Online Purchase | 95.8 | 96.7 |
| Ever Made an Online Sale | 33.4 | 44.6 |

Table 2.7: Students' Stated Preferences

| Stated Preferences | N | Mean | Std. Dev. | Min | Max |
|---|------|------|-----------|-----|-----|
| New Rather than a Used Textbook | 6033 | 1.77 | 2.25 | 0 | 10 |
| Textbook without Notes, etc. | 6033 | 4.00 | 3.09 | 0 | 10 |
| Seeing the Textbook Before Purchasing | 6033 | 3.61 | 2.81 | 0 | 10 |
| Receiving the Textbook on the Day of Purchase | 6033 | 3.20 | 2.55 | 0 | 10 |
| Confidence in Online Shopping | 6033 | 7.36 | 1.81 | 0 | 10 |

Table 2.8: Correlation in Students' Stated Preferences

| | New vs. Used | No Notes | Seeing the Book | Same Day | Online Confidence | Fresh. Dummy |
|-----------------|-----------------|-----------------|--------------------|-----------------|----------------------|-----------------|
| New vs Used | 1.00 | - | - | - | - | - |
| No Notes | 0.51 (0.00) | 1.00 | - | - | - | - |
| Seeing the Book | 0.41 (0.00) | 0.44 (0.00) | 1.00 | - | - | - |
| Same Day | 0.23 (0.00) | 0.20 (0.00) | 0.41 (0.00) | 1.00 | - | - |
| Online Conf. | -0.06 (0.00) | -0.05 (0.00) | -0.20 (0.00) | -0.20 (0.00) | 1.00 | - |
| Fresh. Dum. | 0.05 (0.00) | 0.07 (0.00) | 0.11 (0.00) | 0.04 (0.00) | -0.09 (0.00) | 1.00 |

Table 2.9: Distribution of Payments for Textbooks

| | Freshmen | | Upperclassmen | |
|----------------------------------|-----------|---------|---------------|---------|
| | Frequency | Percent | Frequency | Percent |
| Self | 446 | 31.4 | 1,677 | 44.9 |
| Another person - Varying Payment | 853 | 60.0 | 1,785 | 47.7 |
| Another person - Fixed Payment | 48 | 03.4 | 86 | 02.3 |
| Scholarship | 74 | 05.2 | 191 | 05.1 |

Table 2.10: Search Cost Estimates

| Covariate | Point Estimate | Standard Error |
|-----------------------|----------------|----------------|
| Constant | 10.890 | (2.632) |
| Online Confidence | -0.894 | (.227) |
| First in College | -0.199 | (.146) |
| Prev. Online Purchase | -5.478 | (1.54) |
| Semesters | -0.937 | (.078) |
| Semesters Squared | 0.0636 | (.002) |
| θ^2 | 5.34 | (.125) |

Notes: N = 6033, Simulation Draws = 50
Bootstrapped Standard Errors in Parentheses: 250 Replications

Table 2.11: Utility Parameter Estimates

| Covariate | Point Estimate | Standard Error |
|----------------------------|----------------|----------------|
| Semesters | -0.286 | (.002) |
| Semesters Squared | 0.016 | (.000) |
| Course in Major | -0.106 | (.019) |
| Spring Indicator | -0.740 | (.017) |
| Preference for New | 0.087 | (.005) |
| Preference for Quality | 0.158 | (.005) |
| α - Out of Pocket | -0.560 | (.010) |
| α - Fixed Payment | -0.204 | (.031) |
| α - Varying Payment | -0.309 | (.007) |
| α - Scholarship | -0.964 | (.008) |
| Continuation Value | 0.234 | (.040) |

Notes: N = 6033, Simulation Draws = 50
Bootstrapped Standard Errors in Parentheses: 250 Replications

Table 2.12: Elasticity Estimates

| | Outside Option | New - Bkstore | Used - Bkstore | New - Online | Used - Online | Rental |
|------------------|-------------------|------------------|-------------------|-----------------|------------------|--------|
| Outside Option | . | 0.09 | 0.08 | 0.06 | 0.09 | 0.04 |
| New - Bookstore | . | -0.30 | 0.08 | 0.05 | 0.06 | 0.03 |
| Used - Bookstore | . | 0.08 | -0.29 | 0.04 | 0.06 | 0.03 |
| New - Online | . | 0.06 | 0.05 | -0.33 | 0.11 | 0.05 |
| Used - Online | . | 0.05 | 0.05 | 0.08 | -0.32 | 0.05 |
| Rental | . | 0.06 | 0.07 | 0.07 | 0.15 | -0.41 |

Notes: N = 4234. Excludes textbook obs. not offered online and new/bkstr. price < \$40.

Unbracketed results denote percentages for first-year and upper-level students.

Brackets denote first-year students. Parentheses denote upper-level students.

Elasticities are calculated by increasing the price of the column alternative by 10%.

Table 2.13: Eighth Semester Search Costs Counterfactual

| | | Baseline (%) | Counterfactual (%) | % Change |
|----------------|-------------|--------------|--------------------|----------|
| Outside Option | First Year: | 11.0 | 9.3 | -15.9 |
| | Upper Year: | 17.8 | 17.5 | -1.5 |
| New–Bookstore | First Year: | 24.1 | 19.8 | -17.8 |
| | Upper Year: | 17.1 | 16.8 | -1.6 |
| Used–Bookstore | First Year: | 22.2 | 18.7 | -15.6 |
| | Upper Year: | 17.7 | 17.4 | -1.7 |
| New–Online | First Year: | 14.5 | 16.7 | 15.3 |
| | Upper Year: | 15.2 | 15.5 | 1.7 |
| Used–Online | First Year: | 20.1 | 23.4 | 16.2 |
| | Upper Year: | 22.8 | 23.2 | 1.7 |
| Rental | First Year: | 8.2 | 9.5 | 16.3 |
| | Upper Year: | 9.4 | 9.6 | 1.7 |

Notes: N = 4234. Excludes individuals with textbooks not offered online and new bookstore price < \$40.

CHAPTER 3

PRICE BELIEFS AND EXPERIENCE

3.1 Introduction

Price dispersion is a feature of many markets and even occurs in markets for homogeneous goods or services (Stigler 1961). One possible reason for the persistence of price dispersion is that consumers have limited information over prices and acquiring information may be costly. In markets with limited information and costly search, an individual may not purchase from the seller with the lowest price if she is unaware of that price. Theoretical models of consumer search incorporate the search decision into a model of consumer demand by assuming that individuals have beliefs about the empirical distribution of prices in the market and must incur a cost to reveal price information from one or more retailers before deciding whether to purchase the good or service (e.g., Reinganum, 1979; Burdett and Judd, 1983). The decision to search depends upon the magnitude of the search costs as well as the individual's subjective beliefs about the distribution of prices. When estimating models of consumer search, researchers may impose assumptions on individuals' beliefs in order to recover estimates of search costs. In this paper, we test the validity of these assumptions using data on the observed distribution of prices for the online textbook market and data on individuals' subjective beliefs about this distribution.

There is a growing literature focusing on the development and estimation of structural models of consumer search. These models have been used to explain observed price dispersion for homogeneous goods (Hortaçsu and Syverson 2004; Hong and Shum 2006), test competing models of consumer search (De los Santos et al. 2012b), and to recover demand estimates in markets where price uncertainty is important (Koulayev 2012; Moraga-González et al. 2009). A critical assumption used in these studies is that consumers have rational expectations, (i.e. the price of a product

is a random variable, but consumers know the parameters that govern the distribution of prices. However, if consumers have biased beliefs about the parameters of the empirical distribution of prices, this will lead to biased estimates of search costs. In particular, if consumers' beliefs about prices are biased upward, the rational expectations assumption will bias search cost estimates upwards and bias price elasticity estimates towards zero (low levels of search can be explained by either high search costs or low expected benefits from search). By comparing subjective beliefs to actual observed price distributions, we are able to test the validity of this assumption.

In addition to testing the validity of the rational expectations assumption, we also investigate the degree to which experienced consumers have more accurate beliefs than their less experienced counterparts. Recent research has supported this idea by incorporating learning into consumer search models.¹ In these models, consumers learn about the parameters of the empirical price distribution within a single purchasing decision through a sequential search process (De los Santos et al. 2012a; Koulayev 2009, 2013). We focus instead on learning across purchasing decisions; in particular we examine the hypothesis that more experienced consumers have acquired information about the empirical price distribution through repeated participation in the market.²

We use data on the empirical distribution of textbook prices from online retailers and consumers' subjective beliefs about this distribution. In order to obtain data on individuals' subjective beliefs, we provide an online questionnaire to 1,224 undergraduate students with multiple textbook purchasing scenarios in order to elicit their beliefs about prices. For each hypothetical textbook purchasing scenario, students are given the price of a textbook from the campus bookstore and are asked about their expectations of the lowest price available from an online retailer. Additional

¹Earlier studies examined learning and search through experimental designs (e.g., Sonnemans, 1998; Einav, 2005)

²The research questions we address in this paper are further supported by research in the labor literature, which uses subjective beliefs about future earnings to explain college major choice (Arcidiacono et al. 2012; Stinebrickner and Stinebrickner 2011; Wiswall and Zafar 2012). These studies show that incorporating students' subjective beliefs leads to significantly different estimates than those obtained under the assumption of rational expectations. In addition to this, Wiswall and Zafar (2012) show that college students' beliefs about future earnings become more consistent with the actual earnings distribution as they proceed through school (i.e., become more familiar with their field).

questions are asked to elicit consumers' beliefs about the variability of the lowest price. For example, if a consumer reports an expected online price of \$100, she is then asked about the likelihood that the actual price is below \$95.

Our results show that inexperienced consumers have price expectations that are significantly greater than the mean of the empirical price distribution for both new and used textbooks. Therefore, we can reject the hypothesis that inexperienced consumers know the parameters of the price distribution for the online textbook market. Individuals with higher levels of experience, measured by the number of prior online textbook purchases, typically have price expectations that are closer to the empirical mean. For used books, individuals tend to underestimate the variation of the empirical distribution, and beliefs about the variation of the price distribution do not appear to become more accurate with experience. Overall, the evidence is consistent with learning, at least for learning about the mean of the price distribution.

The following section provides theoretical motivation for this project and expands on our goals. Section 3 describes the data and Section 4 presents results. Section 5 discusses the issue of selection, and section 6 concludes.

3.2 Theoretical Motivation

We use the following simple model of consumer search to motivate the empirical section of this paper. Individuals can purchase a given product from two locations. Assume for simplicity that the search cost is zero for one of the locations, so the individual knows the price of the product at this location. The price of the product at the other location is unknown by the individual, and there is a cost associated with determining this price. Denote the price at the zero search cost location as p^* and the price at the location with a search cost as p , which is a random variable with cumulative density function, $F(p)$. The individual can either purchase the product from the first location or pay some cost, c , to search and discover the price at the other location. If the individual decides to search, he does not incur an additional search cost should he choose to purchase the product from the first location (i.e., search with recall).

The decision rule for the search problem is given by Equation (3.1). An individual chooses to

search if,

$$\int_0^{p^*} (p^* - p) d\tilde{F}_i(p) > c_i \quad (3.1)$$

where \tilde{F}_i denotes an individual's beliefs about the empirical price distribution. The LHS of Equation (3.1) is the expected benefit of search. A individual integrates over the difference between the known (p^*) and unknown price (p), given his beliefs about the distribution of the unknown price. The domain of integration is bounded above by p^* because an individual can costlessly revisit the first location (i.e., the benefit from search is weakly positive).

The RHS of Equation (3.1) is an individual specific search cost c_i . The majority of the structural consumer search literature attempts to recover the distribution of individuals' search costs. In order to do so, the econometrician must make assumptions regarding individuals' beliefs, \tilde{F} . A common assumption regarding individuals' beliefs is that there is no learning over the parameters of the distribution, and individuals have rational expectations. In other words, individuals are assumed to know the parameters of the distribution of p .³

In this paper, we focus on the first two moments of individuals' beliefs. Determining if these moments match the corresponding moments of the empirical price distribution is important for the estimation of search costs. If consumers overestimate the mean of the empirical price distribution, then the model will generate an upwardly biased distribution of search costs under the rational expectations assumption. Similarly, if consumers underestimate the variance of the empirical price distribution, search cost estimates will also be biased upward.⁴

³An alternative to making a parametric assumption on the empirical price distribution and consumers' beliefs is to instead assume that consumers form beliefs non-parametrically based on the empirical CDF of observed prices:

$$F(p) = \tilde{F}_i(p) = \frac{1}{N} \sum_{k=1}^N I[p_k < p]$$

where N is the number of observed prices. If consumers' beliefs are biased relative to the empirical distribution, this leads to similar biases in search costs that are discussed under the assumption of a parametric distribution for prices and beliefs.

⁴Misspecification of beliefs also leads to biases in price elasticity estimates. If individuals' beliefs are biased such that they underestimate the benefit of search (relative to the assumed, true benefit), then the model will recover price

An alternative to rational expectations is to allow uncertainty and learning over the parameters of the price distribution. When individuals search and observe a price draw, they can use this information to update their beliefs according to a learning process (e.g., Bayesian). Even in the learning framework, however, some variant of the rational expectations assumption is commonly used to restrict individuals' initial prior beliefs as the initial priors are typically not separately identified. In the empirical section of the paper we test whether inexperienced individuals have biased beliefs about the parameters of the price distribution. We also examine whether individuals' beliefs are consistent with learning by testing whether more experienced individuals have beliefs that are closer to the parameters of the empirical price distribution.

3.3 Data

We collected data on subjective beliefs about the distribution of prices in the online textbook market through online questionnaires sent to students at the University of North Carolina at Chapel Hill (UNC).⁵ The questionnaires asked individuals about their previous textbook purchasing behavior and presented them with hypothetical textbook purchasing scenarios. We supplement the responses to these textbook purchasing scenarios with price data scraped from an online marketplace for a large number of textbooks. Before providing a summary of both datasets, we will provide more detailed information about the textbook purchasing scenarios.

3.3.1 Textbook Purchasing Scenarios

Each questionnaire contained three randomly assigned hypothetical textbook purchasing scenarios from a total of twelve potential scenarios.⁶ Figure 3.1 is a screenshot of the information provided in one particular scenario.⁷

elasticities that are lower relative to the true elasticities.

⁵Appendix Section A.1.1 contains the text from the online questionnaire provided to consumers. Individuals who agreed to participate in the survey were sent a link to the questionnaire.

⁶These twelve textbooks include four textbooks each from physical sciences, social sciences, and humanities. Of the four textbooks within these general fields, two are from introductory level courses. More information on the characteristics of the textbooks used in the hypothetical purchasing scenarios can be found in Appendix A.1.2.

⁷For each scenario, we randomly assigned respondents to a full information case (title, author, publisher, picture, etc.) or a limited information case. As opposed to the full information case, as seen in Figure 3.1, the limited

After being presented with information about the scenario, respondents were provided with the (actual) price of a new copy of the textbook from the campus bookstore, and were asked to give their expectations about the lowest price they would find for a new copy of this textbook if they searched only one online retailer.⁸ Respondents were then presented with the actual price of a *used* copy of the textbook from the campus bookstore (including taxes) and were asked to give their expectations about the lowest price they would find for a used copy online if they searched one online retailer (including shipping fees).

In order to elicit information about individuals' beliefs about the higher order moments of the price distribution, we then asked respondents for the probability that the price realized after search would be less than X% or greater than Y% of their reported expected price for both new and used copies of the textbook. For example, in Figure 3.1, the new price of the textbook at the campus bookstore for the Fall 2012 semester was \$87.00. If the respondent reported that her expectation of the lowest price for a new copy of the textbook from one online retailer was \$50.00, then the next questions would ask her the probability that the price would be less than \$45.00 and the probability that the price would be greater than \$55.00. In practice, X was randomly drawn from {85, 90, 95} and Y was randomly drawn from {105, 110, 115}.

Given that individuals may not be accustomed to thinking about prices in a probabilistic manner, we first presented individuals with an example in order to help clarify the questions within the textbook purchasing scenarios. In the example, we asked individuals to consider the lowest price they might find for a pair of jeans if they searched one retailer at the mall. This example contained information about probabilities (e.g., that their response should be between 0 and 100 percent) and clarification about the nature of price uncertainty (i.e. that although their best guess might be \$20, there is some chance that the price is actually lower or higher than \$20).

information case only provided information on the title, author, and course.

⁸The bookstore price provided to students explicitly included sales tax. Respondents were asked to include shipping fees when providing their expectation of the lowest price available. Respondents were also reminded to not actually search for the lowest price of the textbook.

3.3.2 Online Questionnaire Data

We conducted two waves of the survey. The first was during the Fall semester of 2012, and the second was during the Spring semester of 2013. For the Fall 2012 and Spring 2013 semesters, 820 and 798 respondents completed the background questions about their previous textbook purchasing experience, respectively. The sample used in analysis is composed of 739 respondents from the Fall 2012 semester and 726 respondents from the Spring 2013 semester.⁹ 104 respondents (52 from both semesters) were dropped because they had been enrolled in college for more than 10 semesters and an additional 49 respondents (29 from Fall 2012 and 20 from Spring 2013) were dropped for reporting nonsensical answers (e.g., reporting an expected price of \$100,000).¹⁰ Appendix A.1.3 provides a more detailed description of within survey attrition.

Table 3.1 displays the number of semesters enrolled for the questionnaire respondents. This count includes both traditional fall and spring semesters and any summer sessions the students had previously been enrolled in. Individuals in later semesters are slightly over sampled due to the nature of how we recruited individuals for the study. We obtained the email addresses of individuals who participated in a separate, longer running data collection project and agreed to receive follow-up emails. This other project began in the Fall of 2011 and recruited new individuals each semester. Individuals who participated at the start of this other project would be at least in their third semester at the time of data collection (assuming continual enrollment). Appendix A.1.1 provides more detail on how individuals were recruited.

Respondents' previous textbook purchasing behavior and major choice are also reported in Table 3.1. A majority of respondents have purchased textbooks at the campus bookstore and from an online retailer. There is significant variation in how many textbooks respondents have purchased online; 33.6% of the individuals in the sample have purchased five textbooks or fewer from online retailers. Approximately a third of respondents reported either Economics or a STEM field as one

⁹There were 240 individuals who participated in both surveys.

¹⁰In practice, this was done by removing respondents who reported expectations less than 10% or greater than 150% of the bookstore price. In Appendix A.2.1 we report out main results for a more relaxed omission criterion. The results are substantially the same.

of their stated majors.

3.3.3 Online Retailer Data

In order to construct an empirical distribution of prices for textbooks, we used a script in Perl to scrape .html files from Amazon.com. We collected daily price data for approximately 3,500 books that were assigned at UNC during the 2012-2013 school year. Using these .html files, we used a separate script in Perl to parse the lowest prices available for both new and used copies of the books on each day.¹¹ Since the survey asked individuals about their expectations of the lowest available price for a particular book, we define the empirical distribution as the distribution of the lowest online price as a proportion of the price at the campus bookstore across textbooks. We use the daily price data for two intervals corresponding to the timing of the surveys. The Fall survey period is from November 30, 2012 to December 10, 2012, and the Spring survey period is from April 11, 2013 to April 26, 2013.¹² To construct the empirical price distribution, we use the average price of the textbook over the survey period. The price sample used in the analysis trims the top and bottom 0.5% of the prices for each survey period.¹³

The total cost of purchasing books online includes shipping fees, which vary depending on the speed of delivery. For items purchased on the Amazon Marketplace from third party sellers, we added the fee for standard shipping. Items purchased directly from Amazon qualify for free standard shipping as long as the item is purchased in as part of an order that exceeds a certain amount.¹⁴ Most new textbooks will qualify for free shipping if purchased directly from Amazon,

¹¹Further analysis could incorporate additional prices from these .htmls files such as the lowest price conditional on reported quality (e.g. very good, good, etc.).

¹²The online questionnaire was initially distributed on November 30, 2012 for the fall survey and April 11, 2013 for the spring survey. Nearly all of the surveys were completed during these intervals. We take these periods as the time frames that individuals are forming their expectations over. This is potentially problematic as online textbook prices vary systematically across the year (e.g., they are generally higher in August than May.). Further analysis could examine additional time frames in the construction of the empirical distribution.

¹³The trimmed sample excludes books that have an online price listing that is either a very small fraction or a large multiple of the bookstore price. In some cases, particularly for books with low sales volumes, the automated pricing algorithms used by larger book resellers can generate these extreme prices.

¹⁴Orders that exceeded \$25 qualified for free shipping at the time of the surveys.

so we do not add any shipping fees to the price of these books.¹⁵ We include sales tax in the campus bookstore prices. Sales taxes are not included in the online prices.¹⁶

Table 3.2 provides the ratio of prices of textbooks from Amazon.com relative to the price from the campus bookstore. The first row reports the prices of new books for the full sample of books for which we have data. On average, new prices on Amazon.com are approximately 85% of the bookstore price. The second to last column reports the average difference between the price of the textbook from the bookstore and an online retailer. For all textbooks in our sample, the savings in absolute terms is approximately \$10.

The second row reports the new prices that includes new books listed on the Amazon marketplace by third party sellers. Including the marketplace listings increases the savings relative to the bookstore price. On average, used prices on Amazon.com are approximately 76% of the used bookstore price. This corresponds to an average difference of approximately \$33. The median is lower than the mean for both new and used books, as the distributions are slightly skewed to the right. On average the prices during the Spring survey period were slightly lower than during the Fall period.

The next three rows of Table 3.2 provide summary statistics for textbooks which are priced greater than \$100 for a new copy from the campus bookstore. Books with a list price below \$100 include popular press titles that have a large market outside of being assigned for a college course. The restricted sample of books which are priced greater than \$100 at the campus bookstore consists primarily of books that are commonly thought of as textbooks. Relative to the full sample, the potential savings from shopping online becomes greater for both new books and used books (i.e. in both percentage and magnitude terms, more expensive textbooks have greater savings

¹⁵We do not include shipping for books that do not individually qualify for free shipping because they could be purchased as part of a larger order that does qualify for free shipping. The empirical analysis focuses on higher priced books that would qualify for free shipping. All of the books in the hypothetical textbook purchasing scenarios qualify for free shipping if purchased new from Amazon.

¹⁶At the time of this analysis, Amazon did not collect sales taxes. Individuals were responsible for paying the sales taxes for online purchases, however compliance was low. Sellers on the Amazon Marketplace are responsible for paying any applicable sales taxes, so sales taxes are already included in the listed prices.

in the online market). The variability of prices is less for both new and used books relative to the full sample. The final three rows provide summary statistics for the textbooks used in the hypothetical textbook purchasing scenarios.¹⁷ On average, these prices are slightly lower than the sample of textbooks with a price of \$100 or more at the campus bookstore, but the difference is not significant.

Ideally, how we define the empirical price distribution should match the price distribution of the individuals' beliefs, but there are a few reasons why this may not be the case. First, textbook prices vary over time, and the time frame used to define the empirical distribution may not match the time frame of the individuals' subjective beliefs. Second, we only use price data from a single online retailer. We believe the prices from Amazon.com provide a reasonable approximation to the empirical distribution of prices that consumers face if they only search one online retailer. Of the individuals in our sample, 75% reported Amazon.com as the first website they would visit to search for a textbook. These issues of timing and alternative retailers affect the comparison between the individuals' subjective beliefs and the empirical price distribution. The comparison of individuals' subjective beliefs across different levels of experience is not affected as long as individuals with different levels of experience do not systematically differ in the time frame considered or in the choice of the online retailer.

3.4 Results

The first subsection presents results using the data on reported expectations. The following subsection incorporates additional data on beliefs to examine not only individuals' expectations but also individuals' beliefs about the variance of the empirical price distribution in the context of a parametric learning model. The online survey asks individuals to report what they thought the price of the textbook would be if they searched one online retailer. We interpret the responses to this question as corresponding to individuals' subjective beliefs about the mean of the price distribution of the lowest price for a particular textbook.

¹⁷Note that the total number of textbooks in the purchasing scenarios is actually 12. However, online retailer data for one textbook is missing.

3.4.1 Expectations Results

In this section we present descriptive statistics of individuals' price expectations. Then, we test for differences in price expectations relative to the empirical prices across levels of experience in order to determine if consumers' expectations converge to the mean of the empirical price distribution. Finally, we perform regressions to control for additional characteristics of the respondents and the textbook scenarios.

The first columns of table 3.3 provide the summary statistics of the reported expectations of the lowest online price as a proportion of the bookstore price for individuals with different levels of online textbook purchasing experience. In the survey, individuals were asked about the number of textbooks they had ever purchased online, and they responded by selecting one of four possible categories. Individuals with no prior online textbook purchases expect the price of a used book online to be approximately 74% of the price of a used book at the college bookstore. This corresponds to an expected savings of \$31.53 on average across the hypothetical textbook purchasing scenarios. Individuals with prior online textbook purchases expect the online price to be lower, with higher levels of experience corresponding with a greater expected savings. On average, individuals with more than ten previous online purchases expect the price of a used book online to be approximately 65% of the price of a used book at the college bookstore. This corresponds to an average savings of \$41.65. The results from the spring survey display a similar pattern.

Table 3.5 repeats the analysis done in table 3.3 using level differences instead of the normalized price ratio. Consumers across experience levels expect for there to be an average savings of \$30 to \$40 for textbooks from online retailers. The patterns across experience levels are the same when using levels as using ratios. As consumers gain experience, they expect to find larger savings in the online market.

These results demonstrate that higher levels of experience are associated with lower expectations of online textbook prices. This relationship would be consistent with learning if the individuals with higher levels of experience report expectations that are closer to the true mean of the price distributions. The final two columns of table 3.3 report the difference between the average of the

reported expectations and the mean of the empirical price distribution for the sample of scenario textbooks as well as the sample of textbooks with a list price greater than \$100. On average, the reported expectations become closer to the empirical mean at higher levels of experience. For the scenario textbooks, the difference between the mean of the reported expectations and the empirical mean is not significant at any level of experience.¹⁸ For the sample of books with list price greater than \$100 for the fall survey, this difference is significant at the lowest levels of experience and is not significant at the higher levels of significance. For the spring survey, the difference between the mean of the reported expectations and the mean of the empirical price distribution is significant at all levels of experience for books with a list price greater than \$100. This is due to the mean of the empirical price distribution being lower during the spring survey period.

Table 3.5 reports the results for new books. Individuals with higher levels of experience report lower expected prices on average. Individuals with no experience expect the online price to be 83% of the price of the textbook from the campus bookstore (a \$26.67 average savings). Individuals with eleven or more previous online purchases expect the online price to be 77% of the price from the campus bookstore (a \$37.58 average savings). The difference between the empirical mean and the mean of the reported expectations decreases for higher levels of experience. However, unlike the results for used books, individuals with higher levels of experience have expectations that are on average significantly below the empirical prices. One explanation for this result is that the new price is the price for purchasing the book directly from Amazon.com. When the new price is defined as the minimum of the marketplace price and the price charged by Amazon.com, the reported expectations are significantly greater than the empirical mean for all levels of experience. Some individuals likely include the marketplace when forming their beliefs about the prices of new textbooks. One possible explanation for the relationship between experience and price expectations for new books is that individuals are learning about the availability of new textbooks by third party sellers.

Due to the nature of the data collection, we want to control for differences in the textbook

¹⁸This result is primarily due to the small sample size for the scenario textbooks.

purchasing scenarios that individuals are given and control for additional characteristics of the individual which may explain the differences in price expectations across levels of experience. Table 3.6 reports results from a regression of normalized price expectations on level of experience, textbook characteristics, scenario characteristics and additional individual controls. The scenario characteristics include indicators to control for the different possible scenarios, the survey period, and whether the textbook purchasing scenario was a full information case (details were provided on textbook characteristics such as years since revision, etc.). The additional individual controls include indicators for whether the individual has previously taken the course for which the textbook was assigned and whether the individual has previously been assigned the textbook in the scenario.

The regression estimates are consistent with the mean comparisons above. Individuals who have never made an online textbook purchase before have significantly higher price expectations than individuals who have purchased a textbook online. Price expectations evolve gradually, as individuals in the highest category of experience consistently have lower expectations.

The coefficients on indicators for whether the respondent had previously taken the course or been assigned the textbook are consistently negative, but only statistically significant for individuals who had previously taken the course. The coefficient on the number of years since the last revision is negative and significant, perhaps reflecting beliefs about a greater supply of textbooks in the secondary market. As the number of years since a textbook has been revised increases, the supply of textbooks in the secondary market increases, generally reducing the price of the textbook. Consumers seem to internalize this when making a textbook purchasing decision, which supports the results in Chevalier and Goolsbee (2009). Similarly, consumers have higher price expectations for textbooks that are the latest edition released (two of the twelve textbook scenarios were for previous editions).

Table 3.7 reports regression results with year in school dummies. These results show that the differences in beliefs are due to differences in direct online textbook purchasing experience rather than from indirect experience (e.g., word of mouth).

3.4.2 Distribution Results

In this subsection, we examine whether the patterns observed in the data are consistent with learning over additional parameters/moments of the empirical price distribution. In the hypothetical textbook purchasing scenarios, individuals report their expectations for prices as well as the probability that a draw from the price distribution is below a given threshold ($\mathbf{E}[H]$ and $F_p(p_L; \mu, \sigma)$). We use these two moments to calculate the expected parameters of each individual's beliefs (i.e., $\mathbf{E}[\mu]$ and $\mathbf{E}[\sigma]$), under the assumption that individuals believe that prices follow a log-normal distribution.¹⁹ The log normal distribution has two properties that make it an appropriate distribution in the current context. First, the support of the distribution is non-negative real numbers and prices are bounded below by zero. The second feature is that the log normal distribution is skewed to the right, which is a feature of both the reported beliefs in the sample and the empirical distribution. The most important criteria is that the beliefs (i.e., prior and posterior distributions) of the distribution parameters are conjugate distributions, which is necessary for tractably modeling a Bayesian learning process. The results are similar under alternative distributional assumptions.²⁰

Assuming that individuals believe that the distribution of prices is log-normal, then individuals' prior distribution on μ and $\frac{1}{\sigma^2}$ is Normal-Gamma. If the individual searches, she observes a price which she uses to update her beliefs. As the number of price observations increases, the individual's mean prior on μ and σ converge to the true parameters, and the variance of the priors converge to zero. In terms of the search problem, evidence of individual learning requires that individuals with more experience in the market (i.e., more observations of prices) have more accurate beliefs about the true parameters of the price distribution and more certainty in their beliefs.

Denote the individual's expected parameters as μ_i and σ_i . In the analysis, we consider the distribution of the individual's expected parameters in the population. Define $\bar{\mu}^e$ and $\bar{\sigma}^e$ as the

¹⁹The parameters of the log-normal distribution this is done by using the following equations for the mean and CDF of a log-normal random variable: $\mathbf{E}[H] = \exp(\mu + \frac{\sigma^2}{2})$, and $F_p(p_L; \mu, \sigma) = \Phi(\frac{\log(p) - \mu}{\sigma})$, where Φ is the standard normal CDF.

²⁰The results for the gamma and normal distributions are presented in the Appendix.

mean of individuals' beliefs with the same level of experience e (i.e. $\bar{\mu}^e = \frac{1}{N^e} \sum_i \mu_i * \mathbf{1}[e_i = e]$ and $\bar{\sigma}^e = \frac{1}{N^e} \sum_i \sigma_i * \mathbf{1}[e_i = e]$). Similarly, define $Var[\mu]^e$ and $Var[\sigma]^e$ as the variance among individuals' beliefs with experience level e . As the number of signals the individual receives increases, the expected parameters should converge to the true price distribution parameters. Since each individual's beliefs converge to the true parameters, $\bar{\mu}^e$ and $\bar{\sigma}^e$ should converge to the true parameters as e increases. The convergence of each individual's beliefs to the true parameters as experience increases implies that the variance among individuals' beliefs goes to zero. However, at low levels of experience, $Var[\mu]^e$ and $Var[\sigma]^e$ may increase depending on the variance among individuals' initial prior beliefs. If individuals have similar initial mean priors, then the signal noise would generate greater dispersion of individuals' beliefs for low levels of experience.

Table 3.8 reports the summary statistics for the reported probability that a draw from the price distribution is below some threshold for different levels of the threshold. The threshold is defined as a fraction of the individual's reported expectation. On average, individuals report that the likelihood of the lowest price being less than 85% of their expected lowest price is 0.298. For higher levels of the threshold, individuals assign a larger probability that the price is below the threshold.

Some individuals report a probability of zero or 100 which cannot be justified given the distributional assumption. Similarly, reported probabilities close to zero or 100 will only fit the distribution for extreme values of the parameters. Once the parameter values are calculated, individuals with parameter values in the top or bottom 2.5% of parameter values for either parameter are dropped from the sample to reduce the impact of outliers.²¹

In order to make the interpretation of the results more straightforward, we use the individual's distribution parameters to calculate the mean and standard deviation of the individual's expected price distribution, which is defined as the distribution with the individual's expected parameter

²¹Probabilities of zero and 100 are replaced with 1 and 99 respectively. The individuals who report probabilities of zero or 100 are included in the 2.5%. For the normal distribution, the initial sample only includes individuals who report a probability less than 50%.

values.²² Table 3.9 reports the sample mean and standard deviation of these moments of the individual's expected price distribution by level of experience. Differences in the mean values from the analysis in the previous section is due to the different samples that result from the different rejection criteria. The mean and standard deviation of the preferred specifications of the empirical distribution are presented for comparison.

For used books, the variability of the mean across individuals with the same level of experience does not decrease for individuals with the highest level of experience. So there is greater variability in the expected lowest price for individuals with the highest level of experience. One reason for the greater variability for the highest category of experience is that there may be greater variability in the underlying level of experience for individuals in this group since it includes a larger range of the number of previous textbook purchases. The mean of the standard deviation of the expected price distribution initially increases with experience (from 0.238 for individuals with no online purchases to 0.250 for individuals with 1 to 5 online purchases) and then decreases with experience for higher levels of experience. The variability of the standard deviation of the expected price distribution across individuals with the same level of experience tends to decrease for higher levels of experience, which is consistent with learning. The significance levels reported for the mean are from a two-sample equality of means test that compares individuals within a particular experience group to everyone not in that group. The test for the equality of variances defines the comparison samples in the same way. For new books, the standard deviation of the expected price distributions and its variability within experience groups display similar patterns as for used books. However, the significance of these patterns is less.

Comparing the beliefs about the standard deviation of the price distribution to the empirical standard deviation suggests that individuals may underestimate the variability of prices for used books and overestimate the variability of prices for new books. For used books, however, individuals' beliefs about the mean of the standard deviation decrease at higher levels of experience,

²²Note that individuals' two responses for each scenario exactly identify their expectations of the mean and variance of the normalized price distribution.

moving farther away from the empirical standard deviation. There are several possible explanations for this result. First, our construction of the empirical price distribution may overstate the variability of prices by including erroneous product listings (e.g., sellers listing old editions or international editions). Also, the empirical distribution we construct may not be representative of the books encountered by the typical student since we include all books that are assigned at the university. Another explanation is the inherent difficulty in eliciting beliefs about variance as individuals may not be accustomed to thinking in probabilistic terms.

Figure 3.2 shows the density function of the log-normal distribution for the mean of the individual parameter values as well as the empirical distribution. Moving from the group with no experience to the group with some experience (1 to 5 online textbook purchases), the price distribution shifts to the left and the variance increases slightly. The distributions for higher levels of experience are similar to the group with some experience but have lower variance. This is in contrast to the empirical distribution, which displays much more variability than the beliefs. Although individuals with experience are more accurate in predicting the mean of the distribution, even high experience individuals tend to place too little weight in the left tail of the price distribution. Figure 3.3 shows the densities for new books. As experience increases, the variance of the distributions decreases. Similar to used books, individuals tend to understate the variability of the empirical distribution but to a lesser degree.

Overall, the evidence is consistent with learning, although the evidence suggests incomplete learning. It may be the case that individuals are only learning over a single parameter. This would explain why individuals with more experience are better able to predict the mean price, but are no better (and are actually worse for used books) in incorporating the variance of the price distribution into their beliefs. Another possibility is that individuals in the sample do not have sufficient experience for the convergence properties of the learning process to be evident.

3.4.3 Price Beliefs by Major

In this section, we test whether there are differences in individuals' beliefs for STEM majors and non-STEM majors. The STEM majors include the natural sciences, math, and other quantitative fields (including Economics). Individuals with multiple majors are categorized as STEM majors if any of their majors are in a STEM field. Table 3.10 reports the average expected price for STEM and non-STEM majors. For both new and used books, there is not a significant difference between the price expectations for individuals with no prior online purchases. For used books, this difference becomes significant at low levels of experience as the price expectations of STEM majors decreases at a faster rate. At higher levels of experience, the price expectations of non-STEM majors appears to "catch up" to the price expectations of STEM majors and the difference is significant at the 10% level. For new books, the difference in price expectations between STEM and non-STEM majors is only significant at the highest level of experience.

Table 3.11 presents the average standard deviation of the expected price distribution for STEM and non-STEM majors by level of experience. STEM majors tend to have lower expectations about the variability of prices and there is little change in the expected price variation across different levels of experience. For non-STEM majors, the variation in the expected price distribution initially increases at the lowest level of experience and decreases at the higher levels of experience. This pattern holds for both new and used books.

The results suggest that individuals in non-STEM majors may incorporate new information about the price distribution differently from STEM majors. The mean of the expected price distribution is higher than the mean of the empirical distribution for individuals with no online purchasing experience regardless of major. If these individuals with no experience search for a textbook online, they are likely to observe a price that is lower than the mean of their expected price distribution. On average, STEM majors incorporate this initial experience by lowering the mean of their expected price distribution while non-STEM majors increase the variance of their expected price distribution.

The results from this section should be interpreted with some caution as there are other factors

that may cause the reported beliefs about the price distribution to differ by major. First, STEM majors may more comfortable answering the kind of probabilistic questions that we ask in the survey. Second, the types of books purchased may be systematically different.

3.5 Learning vs. Selection

Although the evidence is consistent with learning, the differences in individuals' beliefs across levels of experience could also result from selection. If individuals have heterogeneous initial prior beliefs, then individuals who believe that the online price is similar to the bookstore price will not search and will not purchase their books online. Then, if the individuals whose initial priors are close to the true distribution are the ones who search and purchase online, the observed difference in beliefs would be the result of selection based on the initial difference in beliefs and not because of learning.

To distinguish between the effects of learning and selection, we examine the individuals who participated in the survey in both the fall and spring semesters. There were 240 individuals who participated in both surveys. Of these individuals, 89 reported an increase in their level of online textbook purchasing experience from the fall to the spring survey. If selection is generating the observed patterns in the data, then the individuals who report an increase in experience in the spring would have lower expected online prices in the fall than the individuals who do not have an increase in experience. Alternatively, in order for the data to be consistent with learning, then individuals who report an increase in experience should be more likely to report different beliefs in the spring, whereas the beliefs of individuals who do not report an increase in experience should be similar in both periods. For the prior online purchase experience measure, we restrict the analysis to the 22 individuals (between 47 and 56 scenarios) who report no experience in the fall survey. Since this measure of experience is an interval, individuals who remain in the same interval for both fall and spring may or may not have gained experience.²³ The inherent limitation of this test is that one period of learning may not generate a significant difference in beliefs for those whose experience increased. Therefore, this test is primarily a test of the hypothesis of no selection.

²³The results of the test are the same if the sample is not limited to individuals who report no experience in the fall.

Table 3.12 reports the mean parameter values for a log-normal distribution of prices for the two groups for both surveys as well as the mean change in parameter values between surveys. The results of the two-sample t-tests comparing each of the mean values between groups are also reported. There is not a significant difference between the mean parameter values of the two groups in the fall semester for both new and used books. The only difference that is significant is the difference in the value of σ for used books in the fall compared to the spring. However, this change is significant for both groups. These results suggest that selection is not the primary cause of the differences in beliefs across experience levels. However, due to the limited sample size, no definitive conclusion can be drawn.

3.6 Conclusion


Although the evidence is consistent with learning, it appears that the learning process is incomplete. Even individuals with the highest levels of experience on average do not fully converge to the empirical distribution. Also, many individuals with high levels of experience have inaccurate beliefs (i.e., the variation across individuals' beliefs does not converge to zero). There are three primary explanations for the persistence of inaccurate beliefs. The first is that the level of experience where this convergence would occur is beyond what we measure in the data. The second is that the beliefs are converging to a distribution other than what is observed during the sample period. For much of the year, the prices of these textbooks online are relatively stable. For a few weeks prior to the start of the semester, prices rise sharply and peak around the first week of the semester. Since individuals are likely to purchase textbooks during this period, the signal that they receive will be from a distribution with a higher mean than what is observed during the sample period. If an individual only ever purchases books online during the first week of the semester (the time when online prices are greatest), then a high experience individual may expect that potential savings online are relatively modest. Finally, there is likely to be some noise in the reported data as individuals may have differed in their interpretation of questions as well as the amount of consideration given to their responses.

One limitation of this analysis is the problem of external validity. Although the online market

for textbooks is comparable to online markets for other goods, the individuals in our sample are not representative of consumers in other online markets. Relative to consumers in other online markets, individuals in our sample are likely to be younger with higher intellectual ability, but they may have less overall experience in online markets. If there are knowledge spillovers across online markets, i.e. if experience in one online market causes individuals to have more accurate beliefs about the prices in other online markets, then the observed bias in the initial beliefs may be more pronounced in the online textbook market, where individuals are likely to have less overall experience in online markets.

In this paper we use a novel dataset to examine subjective price beliefs and their relationship with experience in a market. We find that inexperienced consumers have biased beliefs, but that consumers appear to be learning about the empirical price distribution as they repeatedly participate in the market. This study also leaves open a wide avenue for future research. First, since we do not estimate a dynamic model of search and learning, we are not able to show how individuals incorporate their beliefs into the search decision. Thus, we are not able to determine whether individuals incorporate the benefits of the additional information obtained through search for future purchasing decisions in their decision to search. Also, if individuals have heterogeneous initial prior beliefs, one potential avenue of future research would be to determine the sources of this heterogeneity. Finally, future research is needed to justify the distributional assumptions on the empirical distribution as well as the prior beliefs.

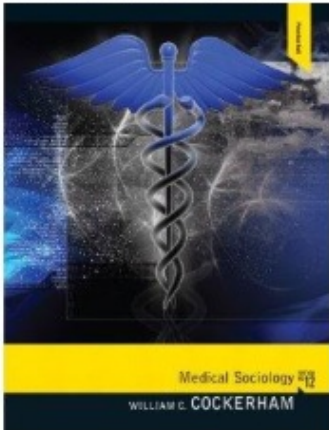
Figure 3.1: Textbook Purchasing Scenario



THE UNIVERSITY
of **NORTH CAROLINA**
at **CHAPEL HILL**

Scenario 1: You are assigned "Medical Sociology" by William Cockerham for an upper-level sociology course (SOCI-469). This is the twelfth and latest edition of the book and was published by Pearson in 2011.

The dimensions of the book are 7 x 0.6 x 9.1 inches, it is a paperback, it contains 432 pages, and weighs 1.2 pounds. A picture is provided below:



Have you ever taken this course?

☐ Yes

☐ No

Have you ever been assigned this textbook?

☐ Yes

☐ No

0% 100%

>>

Table 3.1: Respondent Characteristics

| | Proportion |
|------------------------------------|------------|
| 1 - 2 Semesters | 0.143 |
| 3 - 4 Semesters | 0.248 |
| 5 - 6 Semesters | 0.242 |
| 7 or More Semesters | 0.369 |
| Ever Purchased at Campus Bookstore | 0.960 |
| Never Purchased Online | 0.106 |
| Purchased 1 - 5 Online | 0.230 |
| Purchased 6 - 10 Online | 0.241 |
| Purchased 11 or More Online | 0.423 |
| STEM Major | 0.280 |
| Economics Major | 0.089 |
| Other Major | 0.631 |
| N = 1465 | |

Table 3.2: Ratio of Amazon Prices to Bookstore Prices by Survey Period

| | | Fall Survey Period | | | | | | |
|----------------------------|--------------------|----------------------|-------|-------|--------|-------|-----------------|------|
| | | Mean Ratio | S.D. | Min | Median | Max | Mean Diff. (\$) | N |
| All Books | New | 0.857 | 0.156 | 0.467 | 0.850 | 1.432 | 9.37 | 2051 |
| | New _{alt} | 0.772 | 0.236 | 0.206 | 0.753 | 2.877 | 20.05 | 2220 |
| | Used | 0.758 | 0.383 | 0.091 | 0.715 | 4.207 | 16.24 | 2129 |
| Bookstore Price > \$100 | New | 0.825 | 0.142 | 0.475 | 0.829 | 1.222 | 28.33 | 405 |
| | New _{alt} | 0.678 | 0.169 | 0.206 | 0.673 | 1.377 | 57.29 | 429 |
| | Used | 0.657 | 0.246 | 0.097 | 0.659 | 1.348 | 46.03 | 390 |
| Scenario Books | New | 0.788 | 0.112 | 0.600 | 0.802 | 0.961 | 31.07 | 11 |
| | New _{alt} | 0.659 | 0.136 | 0.514 | 0.609 | 0.861 | 61.48 | 11 |
| | Used | 0.609 | 0.241 | 0.151 | 0.578 | 0.979 | 45.81 | 11 |
| | | Spring Survey Period | | | | | | |
| | | Mean Ratio | S.D. | Min | Median | Max | Mean Diff. (\$) | N |
| All Books | New | 0.834 | 0.148 | 0.357 | 0.838 | 1.425 | 10.99 | 2023 |
| | New _{alt} | 0.755 | 0.295 | 0.202 | 0.737 | 3.970 | 21.64 | 2248 |
| | Used | 0.735 | 0.441 | 0.080 | 0.684 | 5.675 | 18.57 | 2161 |
| Bookstore Price > \$100 | New | 0.795 | 0.158 | 0.358 | 0.815 | 1.326 | 34.02 | 379 |
| | New _{alt} | 0.646 | 0.219 | 0.216 | 0.658 | 1.804 | 62.39 | 434 |
| | Used | 0.597 | 0.279 | 0.085 | 0.602 | 1.731 | 54.57 | 390 |
| Scenario Books | New | 0.729 | 0.120 | 0.524 | 0.768 | 0.879 | 46.57 | 11 |
| | New _{alt} | 0.607 | 0.169 | 0.275 | 0.593 | 0.853 | 66.80 | 11 |
| | Used | 0.614 | 0.255 | 0.123 | 0.581 | 0.919 | 50.26 | 11 |

Notes: The ratio reported is the lowest price on Amazon.com divided by the price of the same title (of equivalent quality) from the campus bookstore. *New_{alt}* refers to the lowest price listed by marketplace sellers for a new copy of the title.

Table 3.3: Mean Ratio Comparisons by Online Purchasing Experience, Used Books

| Fall | | | | | | |
|-----------------------|-------------------------------|------------|-------|--------|-----------------------|---------------|
| Experience | Expectation / Bookstore Price | | | | Mean Expectation Bias | |
| | N | Mean Ratio | S.D. | Median | Scenario Books | Books > \$100 |
| No online purchases | 256 | 0.735 | 0.204 | 0.750 | 0.126 | 0.078*** |
| 1-5 online purchases | 478 | 0.714 | 0.191 | 0.744 | 0.105 | 0.057*** |
| 6-10 online purchases | 477 | 0.663 | 0.172 | 0.683 | 0.054 | 0.006 |
| 11+ online purchases | 810 | 0.645 | 0.184 | 0.645 | 0.036 | -0.012 |
| Spring | | | | | | |
| Experience | Expectation / Bookstore Price | | | | Mean Expectation Bias | |
| | N | Mean Ratio | S.D. | Median | Scenario Books | Books > \$100 |
| No online purchases | 182 | 0.744 | 0.198 | 0.761 | 0.130 | 0.147*** |
| 1-5 online purchases | 439 | 0.710 | 0.170 | 0.745 | 0.097 | 0.113*** |
| 6-10 online purchases | 480 | 0.703 | 0.173 | 0.741 | 0.089 | 0.106*** |
| 11+ online purchases | 888 | 0.660 | 0.176 | 0.675 | 0.046 | 0.063*** |

Notes: The ratio reported is an individual's expectation of the lowest price from an online retailer divided by the price of the same title from the campus bookstore. Expectation Bias refers to the difference between this ratio and the ratio of the observed online price to the bookstore price.

* refers to t-test p-values < .1; ** < .05; *** < .01; H_0 = No difference between ratios.

Table 3.4: Mean Difference Comparisons by Online Purchasing Experience, Used Books

| Fall | | | | | | |
|-----------------------|----------------------------------|------------|-------|--------|-----------------------|---------------|
| Experience | Bookstore Price - Expected Price | | | | Mean Expectation Bias | |
| | N | Mean Diff. | S.D. | Median | Scenario Books | Books > \$100 |
| No online purchases | 256 | 31.55 | 29.86 | 28 | 14.26 | 14.48*** |
| 1-5 online purchases | 478 | 35.02 | 30.51 | 27.5 | 10.79 | 11.01*** |
| 6-10 online purchases | 477 | 40.37 | 30.79 | 31 | 5.44 | 5.66 |
| 11+ online purchases | 810 | 41.73 | 30.78 | 31 | 4.08 | 4.30 |
| Spring | | | | | | |
| Experience | Bookstore Price - Expected Price | | | | Mean Expectation Bias | |
| | N | Mean Diff. | S.D. | Median | Scenario Books | Books > \$100 |
| No online purchases | 182 | 32.16 | 29.52 | 27.5 | 18.10 | 22.41 *** |
| 1-5 online purchases | 439 | 35.22 | 28.56 | 28 | 15.04 | 19.33 *** |
| 6-10 online purchases | 480 | 35.97 | 27.75 | 28.5 | 14.29 | 18.59 *** |
| 11+ online purchases | 888 | 41.28 | 31.09 | 31 | 8.98 | 13.29 *** |

Notes: The difference reported is an individual's expectation of the lowest price from an online retailer subtracted from the price of the same title from the campus bookstore. Expectation Bias refers to the difference between this difference and the difference of the observed bookstore price to the online price.

* refers to t-test p-values < .1; ** < .05; *** < .01; H_0 = No difference between differences.

Table 3.5: Mean Ratio Comparisons by Online Purchasing Experience, New Books

| Fall | | | | | | | | |
|-----------------------|-------------------------------|------------|-------|--------|-----------------------|--------------------|---------------|--------------------|
| Experience | Expectation / Bookstore Price | | | | Mean Expectation Bias | | | |
| | N | Mean Ratio | S.D. | Median | Scenario Books | | Books > \$100 | |
| | | | | | New | New _{alt} | New | New _{alt} |
| No online purchases | 256 | 0.834 | 0.164 | 0.853 | 0.046 | 0.174*** | 0.008 | 0.156*** |
| 1-5 online purchases | 479 | 0.819 | 0.155 | 0.851 | 0.032 | 0.160*** | -0.006 | 0.142*** |
| 6-10 online purchases | 480 | 0.778 | 0.171 | 0.817 | 0.010 | 0.119** | -0.048*** | 0.100*** |
| 11+ online purchases | 814 | 0.768 | 0.159 | 0.798 | -0.020 | 0.109** | -0.058*** | 0.090*** |

| Spring | | | | | | | | |
|-----------------------|-------------------------------|------------|-------|--------|-----------------------|--------------------|---------------|--------------------|
| Experience | Expectation / Bookstore Price | | | | Mean Expectation Bias | | | |
| | N | Mean Ratio | S.D. | Median | Scenario Books | | Books > \$100 | |
| | | | | | New | New _{alt} | New | New _{alt} |
| No online purchases | 184 | 0.835 | 0.183 | 0.870 | 0.103** | 0.225*** | 0.037** | 0.186*** |
| 1-5 online purchases | 444 | 0.818 | 0.150 | 0.856 | 0.089** | 0.210*** | 0.023** | 0.171*** |
| 6-10 online purchases | 486 | 0.788 | 0.152 | 0.822 | 0.059 | 0.180*** | -0.008 | 0.141*** |
| 11+ online purchases | 892 | 0.772 | 0.155 | 0.795 | 0.043 | 0.164*** | -0.023** | 0.126*** |

Notes: The ratio reported is an individual's expectation of the lowest price from an online retailer divided by the price of the same title from the campus bookstore. Expectation Bias refers to the difference between this ratio and the ratio of the observed online price to the bookstore price.

* refers to t-test p-value < .1; ** < .05; *** < .01; H_0 = No difference between ratios.

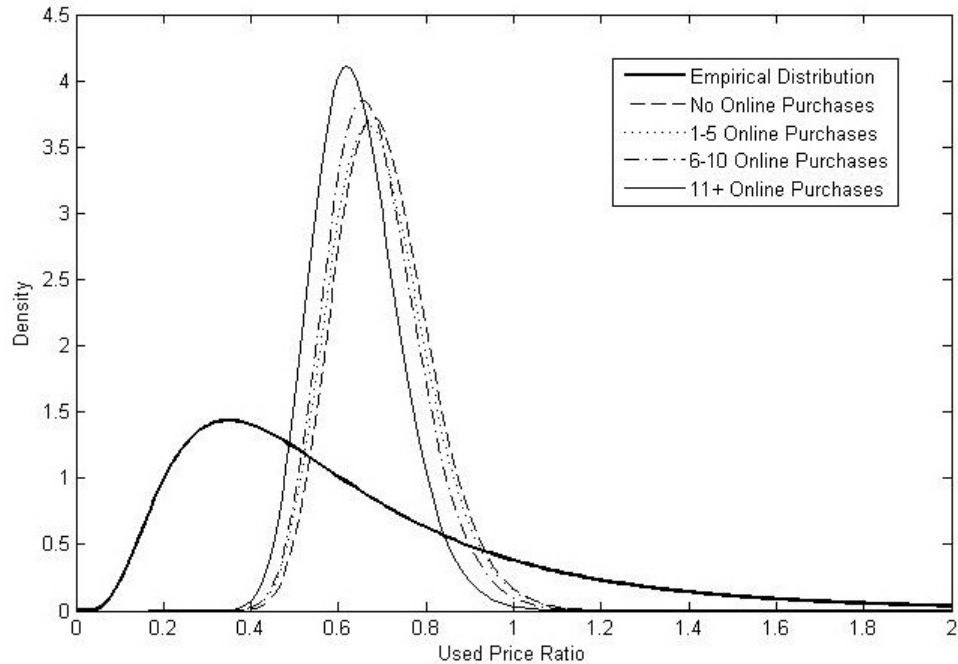


Figure 3.2: Used price pdf versus empirical dist. by level of experience

Table 3.6: (Online Expectation / Bookstore Price) Regressed on Prev. Purchases

| | (1) | | (2) | |
|--|----------------------|----------------------|----------------------|----------------------|
| | New | Used | New | Used |
| 1 - 5 Online Purchases | -0.014 (0.015) | -0.026 (0.017) | -0.015 (0.014) | -0.028* (0.017) |
| 6 - 10 Online Purchases | -0.051*** (0.015) | -0.058*** (0.017) | -0.052*** (0.014) | -0.059*** (0.017) |
| 11+ Online Purchases | -0.063*** (0.014) | -0.087*** (0.016) | -0.063*** (0.014) | -0.088*** (0.016) |
| Previously Taken Course | -0.037*** (0.012) | -0.027** (0.013) | -0.022* (0.012) | -0.025* (0.013) |
| Previously Assigned Book | 0.006 (0.016) | -0.005 (0.017) | -0.008 (0.016) | -0.017 (0.017) |
| Introductory Course | -0.009 (0.006) | 0.003 (0.008) | . | . |
| Latest Edition | 0.007 (0.012) | 0.028** (0.012) | -0.006 (0.015) | -0.003 (0.016) |
| Years Since Last Revision | -0.001* (0.001) | -0.004*** (0.001) | -0.002* (0.001) | -0.002** (0.001) |
| Hardback | -0.009 (0.010) | -0.005 (0.011) | 0.004 (0.011) | -0.016 (0.013) |
| Book Fixed Effects | No | | Yes | |
| Notes: Clustered standard errors (on the individual) given in parentheses. Also included: full information indicator, Spring indicator, pages, and weight. | | | | |
| * refers to p-value < .1; ** < .05; *** < .01 | | | | |

Table 3.7: (Online Expectation / Bookstore Price) Regressed on Experience

| | (1) | | (2) | |
|--|----------------------|----------------------|----------------------|----------------------|
| | New | Used | New | Used |
| 1 - 5 Online Purchases | -0.019 (0.014) | -0.026 (0.017) | -0.020 (0.014) | -0.027 (0.017) |
| 6 - 10 Online Purchases | -0.058*** (0.015) | -0.054*** (0.016) | -0.060*** (0.014) | -0.058*** (0.016) |
| 11+ Online Purchases | -0.076*** (0.014) | -0.084*** (0.016) | -0.077*** (0.014) | -0.087*** (0.016) |
| Second Year | 0.018 (0.013) | -0.018 (0.015) | 0.019 (0.013) | -0.013 (0.015) |
| Third Year | 0.026* (0.013) | -0.012 (0.015) | 0.026* (0.014) | -0.007 (0.015) |
| Four and Above | 0.040*** (0.016) | -0.005 (0.015) | 0.040*** (0.013) | -0.002 (0.015) |
| Additional Controls | No | | Yes | |
| Notes: Clustered standard errors (on the individual) given in parentheses. Additional controls include scenario f.e.s, full information ind., Spring indicator, pages, weight, prev. taken, prev. assigned, latest edition ind., hardback ind. and years since revision. Second year denotes an indicator for individuals in their 3 rd or 4 th semester, etc. | | | | |
| * refers to p-value < .1; ** < .05; *** < .01 | | | | |

Table 3.8: Reported Probability that Lowest Price < $b * \text{Expected Lowest Price}$

| Used | | | | | | |
|------|------|-------|-------|-----|--------|------|
| b | N | Mean | S.D. | Min | Median | Max |
| 0.85 | 1352 | 0.298 | 0.187 | 0 | 0.250 | 0.95 |
| 0.90 | 1368 | 0.328 | 0.189 | 0 | 0.300 | 1.00 |
| 0.95 | 1283 | 0.359 | 0.198 | 0 | 0.300 | 1.00 |
| New | | | | | | |
| b | N | Mean | S.D. | Min | Median | Max |
| 0.85 | 1359 | 0.271 | 0.176 | 0 | 0.250 | 1.00 |
| 0.90 | 1376 | 0.312 | 0.190 | 0 | 0.300 | 1.00 |
| 0.95 | 1293 | 0.339 | 0.195 | 0 | 0.300 | 1.00 |

Table 3.9: Mean and Variance Comparisons (Log-Normal Assumption)

| Used | | | | | |
|-------------------------|------|---------------|---------------|-------------------------------|-------------------------------|
| Empirical Distribution | | | | | |
| | N | Mean | S.D. | | |
| $p_{Bookstore} > \$100$ | 390 | 0.657 | 0.246 | | |
| Scenario Books | 11 | 0.609 | 0.241 | | |
| Beliefs | | | | | |
| Experience | N | Mean $E_i(p)$ | S.D. $E_i(p)$ | Mean $\sqrt{\text{Var}_i(p)}$ | S.D. $\sqrt{\text{Var}_i(p)}$ |
| No online purchases | 370 | 0.713*** | 0.166* | 0.238 | 0.384 |
| 1-5 online purchases | 819 | 0.703*** | 0.166 | 0.250** | 0.362** |
| 6-10 online purchases | 884 | 0.682 | 0.167* | 0.230 | 0.332 |
| 11+ online purchases | 1577 | 0.652*** | 0.171*** | 0.204*** | 0.284*** |
| New | | | | | |
| Empirical Distribution | | | | | |
| | N | Mean | S.D. | | |
| Bkstr. Price > \$100 | 405 | 0.825 | 0.142 | | |
| Scenario Books | 11 | 0.788 | 0.112 | | |
| Beliefs | | | | | |
| Experience | N | Mean $E_i(p)$ | S.D. $E_i(p)$ | Mean $\sqrt{\text{Var}_i(p)}$ | S.D. $\sqrt{\text{Var}_i(p)}$ |
| No online purchases | 371 | 0.813*** | 0.151 | 0.232 | 0.300 |
| 1-5 online purchases | 815 | 0.809*** | 0.145** | 0.246** | 0.346* |
| 6-10 online purchases | 872 | 0.783 | 0.151 | 0.222 | 0.304 |
| 11+ online purchases | 1535 | 0.769*** | 0.149 | 0.215* | 0.289 |

Notes: The significance levels reported for the mean values are from a two-sample equality of means test. The significance levels for the standard deviations are from Brown and Forsythe’s alternative formulation of Levene’s robust two-sample equality of variances test.

* refers to p-value < .1; ** < .05; *** < .01

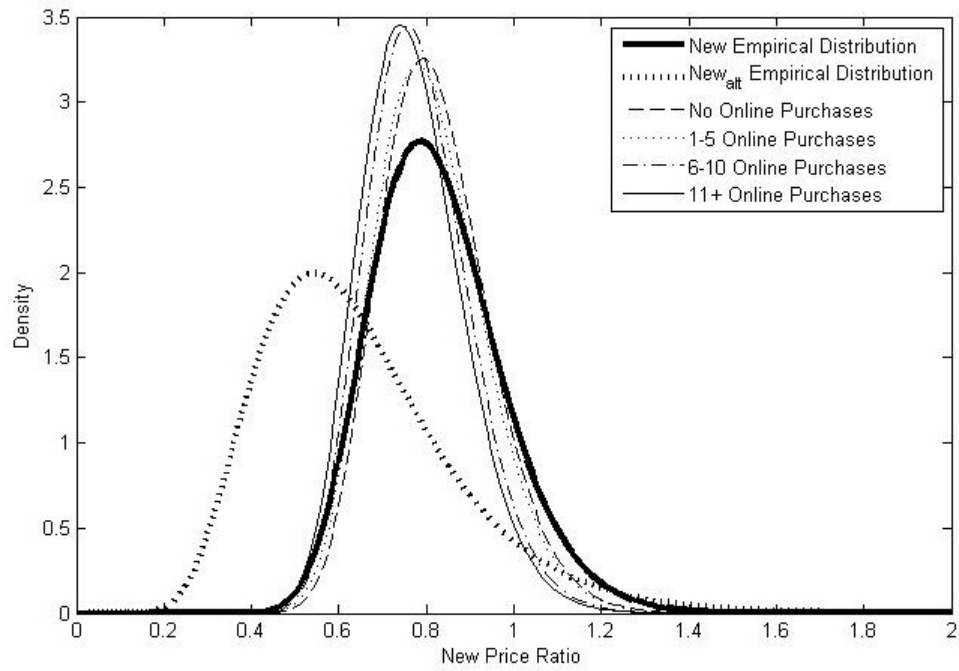


Figure 3.3: New price pdf versus empirical dist. by level of experience
 New_{alt} is the minimum price for a new textbook from Amazon or Amazon Marketplace.

Table 3.10: Mean Comparison by Major

| Used Books | | | | | |
|-----------------------|-------------|---------------|-----------------|---------------|---------|
| Experience | STEM majors | | Non-STEM majors | | p-value |
| | N | Mean $E_i(p)$ | N | Mean $E_i(p)$ | |
| No online purchases | 117 | 0.703 (0.167) | 252 | 0.718 (0.166) | 0.429 |
| 1-5 online purchases | 314 | 0.681 (0.174) | 505 | 0.716 (0.160) | 0.004 |
| 6-10 online purchases | 323 | 0.667 (0.175) | 560 | 0.692 (0.161) | 0.036 |
| 11+ online purchases | 598 | 0.642 (0.171) | 979 | 0.657 (0.171) | 0.093 |
| New Books | | | | | |
| No online purchases | 118 | 0.820 (0.143) | 259 | 0.811 (0.154) | 0.568 |
| 1-5 online purchases | 320 | 0.801 (0.149) | 508 | 0.814 (0.144) | 0.245 |
| 6-10 online purchases | 321 | 0.787 (0.150) | 563 | 0.779 (0.155) | 0.426 |
| 11+ online purchases | 596 | 0.754 (0.148) | 966 | 0.777 (0.152) | 0.003 |

Notes: The p-value is from a two-sample equality of means test.

Standard deviation in parentheses.

Table 3.11: Variance Comparison by Major (Log-Normal Assumption)

| Used Books | | | | | |
|-----------------------|-------------|---------------------------------|-----------------|---------------------------------|---------|
| Experience | STEM majors | | Non-STEM majors | | p-value |
| | N | Mean $\sqrt{\mathbf{Var}_i(p)}$ | N | Mean $\sqrt{\mathbf{Var}_i(p)}$ | |
| No online purchases | 117 | 0.220 (0.387) | 252 | 0.248 (0.384) | 0.528 |
| 1-5 online purchases | 314 | 0.207 (0.224) | 505 | 0.276 (0.424) | 0.003 |
| 6-10 online purchases | 323 | 0.215 (0.292) | 560 | 0.240 (0.353) | 0.249 |
| 11+ online purchases | 598 | 0.209 (0.292) | 979 | 0.202 (0.279) | 0.642 |
| New Books | | | | | |
| No online purchases | 118 | 0.203 (0.290) | 259 | 0.252 (0.336) | 0.153 |
| 1-5 online purchases | 320 | 0.213 (0.292) | 508 | 0.278 (0.421) | 0.009 |
| 6-10 online purchases | 321 | 0.231 (0.326) | 563 | 0.222 (0.322) | 0.699 |
| 11+ online purchases | 596 | 0.201 (0.270) | 966 | 0.225 (0.316) | 0.114 |

Notes: The p-value is from a two-sample equality of means test.

Standard deviation in parentheses.

| Used Books | | | | | | | | |
|---------------|------|----------------|-----------------|--------|----------------|-----------------|--------------|-----------------|
| Group | Fall | | | Spring | | | Difference | |
| | N | mean μ_i | mean σ_i | N | mean μ_i | mean σ_i | mean μ_i | mean σ_i |
| Increase Exp. | 48 | -0.447 (0.321) | 0.297 (0.307) | 47 | -0.412 (0.356) | 0.199 (0.169) | 0.034 | -0.098* |
| Same Exp. | 55 | -0.479 (0.363) | 0.342 (0.372) | 54 | -0.446 (0.299) | 0.230 (0.255) | 0.034 | -0.111* |
| Difference | | 0.033 | -0.045 | | 0.034 | -0.032 | | |
| New Books | | | | | | | | |
| Group | Fall | | | Spring | | | Difference | |
| | N | mean μ_i | mean σ_i | N | mean μ_i | mean σ_i | mean μ_i | mean σ_i |
| Increase Exp. | 45 | -0.243 (0.163) | 0.216 (0.148) | 50 | -0.294 (0.274) | 0.240 (0.276) | 0.051 | -0.024 |
| Same Exp. | 56 | -0.302 (0.289) | 0.277 (0.279) | 53 | -0.288 (0.217) | 0.246 (0.221) | -0.014 | 0.031 |
| Difference | | 0.059 | -0.060 | | -0.005 | -0.006 | | |

Table 3.12: Parameter Values by Change in Experience

APPENDIX A

APPENDIX TO CHAPTER 2: EXPERIENCE AND THE VALUE OF SEARCH

A.1 Estimation

This section first provides a detailed description of the estimation procedure for the two demand model variations: the structural search model and consideration set approach.

A.1.1 Structural Search Model Estimation

Recall that an individual chooses to search if,

$$\int_{U_{iL}^*}^{\infty} (U_{iO}^* - U_{iL}^*) dG(U_{iO}^*) > c_i \quad (\text{A.1})$$

where $U_{iL}^* = \max\{u_{i0}, u_{i1}, u_{i2}, u_{i3}\}$ and $U_{iO}^* = \max\{u_{i4}, u_{i5}, u_{i6}\}$ denote the maximal utilities in the limited (Ω_L) and online choice set cases (Ω_O). The function G represents individuals' beliefs about the distribution of the maximal utility from the online alternatives. This reflects uncertainty over unobservable demand shocks and the prices of the online alternatives. It is assumed that the unobservable demand shocks follow a Type 1 Extreme Value Distribution¹. In the future I plan to relax this assumption by incorporating a variation of the generalized nesting structure used in Bresnahan et al. (1997).

$$c_i \sim \ln \mathcal{N}(\gamma' W_i, \theta^2) \quad (\text{A.2})$$

Calculating $\int_{U_{iL}^*}^{\infty} (U_{iO}^* - U_{iL}^*) dG(U_{iO}^*)$

When individuals make the search decision, they face uncertainty over the price of the online alternatives and unobservable factors that shift demand after the search decision is made. First

¹Ignoring uncertainty over prices for the moment, recall that the CDF of the maximal element of multiple iid Type 1 Extreme Value random variables drawn from a choice set Ω_O is given by the following:

$$P(U_{iO}^* < x) = G(x; \Omega_O) = \exp\left(- \sum_{j \in \Omega_O} \exp(v_{ij} - x)\right)$$

where v_{ij} represents the observable (to the econometrician) portion of the alternative specific utility function ($u_{ij} = v_{ij} + \epsilon_{ij}$).

consider the case where prices of the online alternatives are known with certainty.

For any value of U_{iL}^* , $\int_{U_{iL}^*}^{\infty} (U_{iO}^* - U_{iL}^*) dG(U_{iO}^*)$ is simulated using the following algorithm:

1. Take NS_1 draws from a standard uniform distribution²: $\mu \sim U(0, 1)$. Each individual draw will be denoted with a superscript s (e.g., μ^s is a scalar draw from standard uniform).
2. Calculate the inverse CDF of U_{iO}^* for each μ^s :

$$G^{-1}(\mu^s) = -\ln\left(\frac{-\ln(\mu^s)}{\sum_{j \in \Omega_O} \exp(v_{ij})}\right) \quad (\text{A.3})$$

For each s , this is essentially taking a draw from U_{iO}^* .

3. If the draw from U_{iO}^* is greater than U_{iL}^* , then assign y^s , the expected benefit of search for one simulation draw, equal to the draw for U_{iO}^* minus the value of U_{iL}^* , otherwise $y^s = 0$:

$$y^s = \begin{cases} G^{-1}(\mu^s) - U_{iL}^* & \text{if } \mu^s > G(U_{iL}^*) \\ 0 & \text{if } \mu^s \leq G(U_{iL}^*) \end{cases}$$

4. The simulated value of $\int_{U_{iL}^*}^{\infty} (U_{iO}^* - U_{iL}^*) dG(U_{iO}^*)$ is given by,

$$\frac{1}{NS_1} \sum_{s=1}^{NS_1} y^s \quad (\text{A.4})$$

Calculating \bar{u}_i

Note that the above subsection took U_{iL}^* as given. However, to the econometrician, U_{iL}^* is a random variable. Because of this, it is helpful to calculate each individual's reservation utility: the value of U_{iL}^* such that the individual is indifferent between searching and not searching. This reservation utility, \bar{u}_i is defined as the value such the following equation holds:

²Draws from standard uniform distributions used for simulation are taken from Halton sequences. Each sequence is formed using a different prime number and the twenty initial draws from each sequence are discarded (to avoid correlation across sequences).

$$\int_{\bar{u}_i}^{\infty} (U_{iO}^* - \bar{u}_i) dG(U_{iO}^*) = c_i \quad (\text{A.5})$$

Note first that the LHS can be simulated for any value of \bar{u}_i using the algorithm in Subsection A.1.1. Note further that this expression is completely independent of the maximal utility an individual *actually* receives from the limited choice set case (U_{iL}^*). From the econometrician's perspective, all that is needed for the calculation of \bar{u}_i is an individual's search cost and the observable portion of the utility they receive from the online alternatives. Each individual's value of \bar{u}_i is estimated by minimizing the following expression with respect to \bar{u}_i :³

$$\left[\int_{\bar{u}_i}^{\infty} (U_{iO}^* - \bar{u}_i) dG(U_{iO}^*) - c_i \right]^2 \quad (\text{A.6})$$

Choice Probabilities

This subsection outlines choice probabilities taking search costs as given and ignoring price uncertainty. These factors will be incorporated in the following section.

The form of the individual likelihoods can be separated into three cases:

1. The likelihood of an individual not searching and choosing the outside option or a bookstore alternative
2. The likelihood of an individual searching and choosing the outside option or a bookstore alternative
3. The likelihood of an individual searching and choosing an online alternative

- The likelihood of an individual not searching and choosing the outside option or a bookstore

³This equation is minimized rather than solving for \bar{u}_i directly in Equation (2.11) because a solution to this equation may not exist for certain values of the utility function and search cost parameters (i.e., for certain parameter values, $c_i > \int_{\bar{u}_i}^{\infty} (U_{iO}^* - \bar{u}_i) dG(U_{iO}^*) \forall \bar{u}_i$).

option, $j \in \{0, 1, 2, 3\}$, is

$$Pr(u_{ij} > u_{ik}, \forall k \neq j \in \Omega_L) \cdot Pr(U_{iL}^* > \bar{u}_i) \quad (\text{A.7})$$

The first term is the probability that the individual chooses the outside option or bookstore alternative conditional on choosing from the limited choice set. The second term is the probability that the maximal utility in the limited choice set is greater than the individual's reservation utility (and thus the individual chooses not to search).

The first term is calculated as,

$$Pr(u_{ij} > u_{im}, \forall m \neq j \in \Omega_L) = \frac{\exp(v_{ij})}{\sum_{k \in \Omega_L} \exp(v_{ik})} \quad (\text{A.8})$$

The second term is calculated as,

$$Pr(U_{iL}^* > \bar{u}_i) = 1 - G(\bar{u}_i; \Omega_L) \quad (\text{A.9})$$

$$= 1 - \exp\left[-\sum_{j \in \Omega_L} \exp(v_{ij} - \bar{u}_i)\right] \quad (\text{A.10})$$

• The likelihood of an individual searching and choosing the outside option or a bookstore option, $j \in \{0, 1, 2, 3\}$ is

$$Pr(u_{ij} > u_{ik}, \forall k \neq j \in \Omega_L) \cdot Pr(U_{iL}^* < \bar{u}_i) \cdot Pr(U_{iO}^* < U_{iL}^* | U_{iL}^* < \bar{u}_i) \quad (\text{A.11})$$

The first term is the probability that the individual chooses the outside option or bookstore alternative conditional on choosing from the limited choice set (Equation A.8). This probability is conditional on the limited choice set rather than the full choice set because the decision to choose an alternative from the limited choice set following search ensures that the maximal utility from the limited choice set is greater than the utility from any of the online alternatives. The second term is the probability that the maximal utility from the limited choice set is less than the individual's

reservation utility (and thus the individual chooses to search). The third term is the probability that the maximal utility from an alternative in the limited choice set is greater than the maximal utility from an online alternative.

The second term is calculated as,

$$Pr(U_{iL}^* < \bar{u}_i) = G(\bar{u}_i; \Omega_L) \quad (\text{A.12})$$

$$= \exp \left[- \sum_{j \in \Omega_L} \exp(v_{ij} - \bar{u}_i) \right] \quad (\text{A.13})$$

The third term must be simulated. The steps are as follows,

1. Take NS_2 draws from a standard uniform distribution: $\mu_2 \sim U(0, 1)$.
2. Calculate an individual's probability of search: $G(\bar{u}_i; \Omega_L)$ (Equation A.13).
3. Truncate the uniform draws so that they lie between 0 and $G(\bar{u}_i; \Omega_L)$

$$\tilde{\mu}_2 \sim U(0, G(\bar{u}_i; \Omega_L)) \quad (\text{A.14})$$

4. Calculate the inverse CDF of U_{iL}^* for each $\tilde{\mu}_2^s$:

$$G^{-1}(\tilde{\mu}_2^s) = -\ln \left(\frac{-\ln(\tilde{\mu}_2^s)}{\sum_{j \in \Omega_L} \exp(v_{ij})} \right) \quad (\text{A.15})$$

This is equivalent to taking a draw from U_{iL}^* conditional on U_{iL}^* being less than \bar{u}_i . Label this U_{iL}^{*s}

5. For each s , calculate the probability that U_{iO}^* is less than U_{iL}^{*s} and label this z^s :

$$z^s = Pr(U_{iO}^* < U_{iL}^{*s}) = G(U_{iL}^{*s}; \Omega_O) \quad (\text{A.16})$$

$$= \exp \left[- \sum_{j \in \Omega_O} \exp(v_{ij} - U_{iL}^{*s}) \right] \quad (\text{A.17})$$

6. Calculate the simulated probability as,

$$Pr(U_{iO}^* < U_{iL}^* | U_{iL}^* < \bar{u}_i) = \frac{1}{NS_2} \sum_{s=1}^{NS_2} z^s \quad (\text{A.18})$$

• The likelihood of an individual searching and choosing an online or rental option, $j \in \{4, 5, 6\}$ is

$$Pr(u_{ij} > u_{ik}, \forall k \neq j \in \Omega_O) \cdot Pr(U_{iL} < \bar{u}_i) \cdot Pr(U_{iO} > U_{iL} | U_{iL} < \bar{u}_i) \quad (\text{A.19})$$

The first term is the probability that the individual chooses an online alternative conditional on choosing from the online choice set. This probability is conditional on the online choice set rather than the full choice set because the decision to choose an online alternative following search ensures that the maximal utility from the online alternatives is greater than the utility from any of the alternatives in the limited choice set. The second term is the probability that the maximal utility from the limited choice set is less than the individual's reservation utility (and thus the individual chooses to search, Equation A.13). The third term is the probability that the maximal utility from an online alternative is greater than the maximal utility from an alternative in the limited choice set (one minus the probability simulated according to Equation A.18).

The first term is calculated as,

$$Pr(u_{ij} > u_{im}, \forall m \neq j \in \Omega_O) = \frac{\exp(v_{ij})}{\sum_{k \in \Omega_O} \exp(v_{ik})} \quad (\text{A.20})$$

Incorporating Random Search Costs and Price Uncertainty

Because search costs are random variables and individuals must integrate over their beliefs for the online prices, the choice probabilities in Subsection A.1.1 need to be integrated over individual's search costs and beliefs about the distribution of prices of the online alternatives.

It is assumed that individual's know the empirical distribution of online prices. This will be

described in greater detail in the future.

Choice probabilities are simulated by calculating choice probabilities for each draw from the price distributions and each draw from the cost distribution and then averaged. The complete algorithm for simulating the choice probabilities is as follows,

1. Take a draw c_i^s and draws from each of the online price distributions p_4^s , p_5^s , and p_6^s for each $s \in NS_3$.
2. Simulate $\int_{U_{iL}^*}^{\infty} (U_{iO}^* - U_{iL}^*) dG(U_{iO}^*)$ for each individual.
3. Estimate \bar{u}_i for each individual through the minimization routine described in Subsection A.1.1.
4. Calculate choice probabilities for each $s \in NS_3$.
5. Average each simulated choice probability for the NS_3 draws. Note that there are eleven potential choice paths (seven if the individual chooses to search and four if the individual does not search⁴). So for $j \in \{1, 11\}$,

$$\tilde{P}_{ij} = \frac{1}{NS_3} \sum_{s=1}^{NS_3} P_{ij} \quad (\text{A.21})$$

Note that there are 3 different steps in which simulation methods are used:

1. Simulating $\int_{U_{iL}^*}^{\infty} (U_{iO}^* - U_{iL}^*) dG(U_{iO}^*)$
2. Simulating $Pr(U_{iO}^* < U_{iL}^* | U_{iL}^* < \bar{u}_i)$

| | | |
|--|------------------------------|----------------------------------|
| ${}^4j \in \left\{ \begin{array}{l} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} \right.$ | 1 Search; Outside option | 8 No Search; Outside option |
| | 2 Search; New / Bookstore | 9 No Search; New / Bookstore |
| | 3 Search; Used / Bookstore | 10 No Search; Used / Bookstore |
| | 4 Search; Rental / Bookstore | 11 No Search; Rental / Bookstore |
| | 5 Search; New / Online | |
| | 6 Search; Used / Online | |
| | 7 Search; Rental / Online | |

3. Integrating over the distribution of search costs and online prices.

The first two loops are done sequentially. However, the third loop nests the first two loops, which significantly increases the time needed for estimation. This is necessary for two reasons. The first is because \bar{u}_i must be calculated for each draw from the search cost and price distributions. In order to calculate this, it is necessary to first do Step 1 above. Step 2 must be performed within Step 3 because for each draw within Step 3, the conditional probability must be simulated in order to calculate the full choice probability.

Likelihood Function

Using the simulated choice probabilities given in Subsection A.1.1, the parameters of the alternative specific utility function and search cost distribution, collectively Θ , can be estimated using simulated maximum likelihood estimation. The individual likelihood function is given by,

$$\mathbb{L}_i(\Theta) = \prod_j \tilde{P}_{ij}^{d_{ij}} \quad (\text{A.22})$$

for $j \in \{1, \dots, 11\}$. \tilde{P}_{ij} are the simulated choice probabilities given in Subsection A.1.1 and d_{ij} is an indicator function equal to one if individual i chooses choice path j .

The full log-likelihood function used in estimation is given by,

$$\mathcal{L}(\Theta) = \sum_i \sum_j d_{ij} \ln(\tilde{P}_{ij}) \quad (\text{A.23})$$

A.1.2 Consideration Set Approach Estimation

The consideration set approach posits that consumers make a two step decision: they first decide whether to search for online options and then, conditional on their choice set, make a product decision. However, instead of estimating a set of utility and search cost parameters that determine both the search and product choice, the consideration set approach estimates two sets of parameters: one set that determines the search decision and another set that determines the product choice decision.

An individual chooses to search if,

$$\Theta' X_i^{search} + \epsilon_i^{search} > \Gamma' W_i \quad (\text{A.24})$$

The second stage is as above,

$$u_{ij}^p = \beta' X_{ij} + \phi_j' Y_i + \alpha_{1i} p_j + \alpha_2 \max\{r_{ij}, \bar{r}_{ij}\} + \epsilon_{ij}^{product} \quad (\text{A.25})$$

where the superscript p is used to refer to the product decision (second stage decision). Note that the second stage has the same form for both the structural search and consideration set approaches, but the estimated parameters vary across approaches.

Distribution of Unobservables

The first stage decision rule can equivalently be rewritten as an individual chooses to search if $u_{i0}^s > u_{i1}^s$, where the latent utility from the decisions to search and not search are as follows:

$$u_{i0}^s = \Theta' X_i^{search} + \Gamma' W_i + \epsilon_{i0}^s \quad (\text{A.26})$$

$$u_{i1}^s = \epsilon_{i1}^s \quad (\text{A.27})$$

$$(\text{A.28})$$

where the s superscript denotes the search decision. For ease of notation I denote the vector of epsilons as ϵ_i^s :

$$\epsilon_i^s = \begin{bmatrix} \epsilon_{i0}^s \\ \epsilon_{i1}^s \end{bmatrix} \quad (\text{A.29})$$

Similarly, the epsilons in the second stage are denoted as ϵ_i^p :

$$\epsilon_i^p = \begin{bmatrix} \epsilon_{i0}^p \\ \epsilon_{i1}^p \\ \epsilon_{i2}^p \\ \epsilon_{i3}^p \\ \epsilon_{i4}^p \\ \epsilon_{i5}^p \\ \epsilon_{i6}^p \end{bmatrix} \quad (\text{A.30})$$

If $\epsilon_i^s \sim \mathcal{N}(0, \Sigma^s)$ and $\epsilon_i^p \sim \mathcal{N}(0, \Sigma^p)$, then the estimation procedure simplifies to separately estimating a binary probit and a multinomial probit. However, this implicitly makes the assumption that unobservable factors affecting the decision to search online retailers are independent of any unobservable demand shocks individuals receive from the online alternatives (or any other alternative). In order to relax this restriction, it is assumed that the unobservables in both equations are drawn from a joint distribution: $\epsilon_i \sim \mathcal{N}(0, \Sigma)$:⁵

$$\epsilon_i = \begin{bmatrix} \epsilon_i^{search} \\ \epsilon_i^{product} \end{bmatrix} = \begin{bmatrix} \epsilon_{i0}^s \\ \epsilon_{i1}^s \\ \epsilon_{i0}^p \\ \epsilon_{i1}^p \\ \epsilon_{i2}^p \\ \epsilon_{i3}^p \\ \epsilon_{i4}^p \\ \epsilon_{i5}^p \\ \epsilon_{i6}^p \end{bmatrix} \sim \mathcal{N}(0, \Sigma) \quad (\text{A.31})$$

⁵The exposition of this section follows closely to portions of Chapter Five of Train (2009). The primary difference is the inclusion of sequential decisions with different choice sets across each stage and different choice sets within the second stage conditional on the first stage decision.

$$\underbrace{\Sigma}_{9 \times 9} = \begin{bmatrix} \sigma_{00}^s & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{10}^s & \sigma_{11}^s & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{00}^{sp} & \sigma_{01}^{sp} & \sigma_{00}^p & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{10}^{sp} & \sigma_{11}^{sp} & \sigma_{10}^p & \sigma_{11}^p & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{20}^{sp} & \sigma_{21}^{sp} & \sigma_{20}^p & \sigma_{21}^p & \sigma_{22}^p & \cdot & \cdot & \cdot & \cdot \\ \sigma_{30}^{sp} & \sigma_{31}^{sp} & \sigma_{30}^p & \sigma_{31}^p & \sigma_{32}^p & \sigma_{33}^p & \cdot & \cdot & \cdot \\ \sigma_{40}^{sp} & \sigma_{41}^{sp} & \sigma_{40}^p & \sigma_{41}^p & \sigma_{42}^p & \sigma_{43}^p & \sigma_{44}^p & \cdot & \cdot \\ \sigma_{50}^{sp} & \sigma_{51}^{sp} & \sigma_{50}^p & \sigma_{51}^p & \sigma_{52}^p & \sigma_{53}^p & \sigma_{54}^p & \sigma_{55}^p & \cdot \\ \sigma_{60}^{sp} & \sigma_{61}^{sp} & \sigma_{60}^p & \sigma_{61}^p & \sigma_{62}^p & \sigma_{63}^p & \sigma_{64}^p & \sigma_{65}^p & \sigma_{66}^p \end{bmatrix} \quad (\text{A.32})$$

Above, σ^{sp} denotes a covariance term between an epsilon in the search decision and an epsilon in the product decision. Clearly, normalizations still need to be made to estimate this model. Just as with any discrete choice model, normalizations need to be made for level and for scale.

In order to normalize for level, the epsilons in the search decision and the epsilons in the product decision are differenced by the “zero” alternative:

$$\tilde{\epsilon}_{i0} = \begin{bmatrix} \epsilon_{i1}^s - \epsilon_{i0}^s \\ \epsilon_{i1}^p - \epsilon_{i0}^p \\ \epsilon_{i2}^p - \epsilon_{i0}^p \\ \epsilon_{i3}^p - \epsilon_{i0}^p \\ \epsilon_{i4}^p - \epsilon_{i0}^p \\ \epsilon_{i5}^p - \epsilon_{i0}^p \\ \epsilon_{i6}^p - \epsilon_{i0}^p \end{bmatrix} = \begin{bmatrix} \tilde{\epsilon}_i^s \\ \tilde{\epsilon}_{i1}^p \\ \tilde{\epsilon}_{i2}^p \\ \tilde{\epsilon}_{i3}^p \\ \tilde{\epsilon}_{i4}^p \\ \tilde{\epsilon}_{i5}^p \\ \tilde{\epsilon}_{i6}^p \end{bmatrix} \sim \mathcal{N}(0, \bar{\Sigma}_0) \quad (\text{A.33})$$

$$\begin{aligned}
\underbrace{\tilde{\Sigma}_0}_{7 \times 7} &= \begin{bmatrix} \sigma_{11}^s + \sigma_{00}^s - 2\sigma_{10}^s & \cdot & \dots & \cdot \\ \sigma_{11}^{sp} - \sigma_{10}^{sp} - \sigma_{01}^{sp} + \sigma_{00}^{sp} & \sigma_{11}^p + \sigma_{00}^p - 2\sigma_{10}^p & \dots & \cdot \\ \vdots & \vdots & \ddots & \cdot \\ \sigma_{61}^{sp} - \sigma_{60}^{sp} - \sigma_{01}^{sp} + \sigma_{00}^{sp} & \cdot & & \sigma_{66}^p + \sigma_{00}^p - 2\sigma_{60}^p \end{bmatrix} \\
&= \begin{bmatrix} \tilde{\sigma}_{11}^s & \cdot & \dots & \cdot \\ \tilde{\sigma}_{11}^{sp} & \tilde{\sigma}_{11}^p & \dots & \cdot \\ \vdots & \vdots & \ddots & \cdot \\ \tilde{\sigma}_{61}^{sp} & \cdot & & \tilde{\sigma}_{66}^p \end{bmatrix}
\end{aligned}$$

In order to normalize for scale, $\tilde{\sigma}_{11}^s$ and $\tilde{\sigma}_{11}^p$ are normalized to equal one. Thus the normalized covariance matrix of epsilons differenced against the zero alternatives, $\tilde{\Sigma}_0$ takes the following form:

$$\tilde{\Sigma} = \begin{bmatrix} 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \tilde{\sigma}_{11}^{sp} & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \tilde{\sigma}_{21}^{sp} & \tilde{\sigma}_{21}^p & \tilde{\sigma}_{22}^p & \cdot & \cdot & \cdot & \cdot \\ \tilde{\sigma}_{31}^{sp} & \tilde{\sigma}_{31}^p & \tilde{\sigma}_{32}^p & \tilde{\sigma}_{33}^p & \cdot & \cdot & \cdot \\ \tilde{\sigma}_{41}^{sp} & \tilde{\sigma}_{41}^p & \tilde{\sigma}_{42}^p & \tilde{\sigma}_{43}^p & \tilde{\sigma}_{44}^p & \cdot & \cdot \\ \tilde{\sigma}_{51}^{sp} & \tilde{\sigma}_{51}^p & \tilde{\sigma}_{52}^p & \tilde{\sigma}_{53}^p & \tilde{\sigma}_{54}^p & \tilde{\sigma}_{55}^p & \cdot \\ \tilde{\sigma}_{61}^{sp} & \tilde{\sigma}_{61}^p & \tilde{\sigma}_{62}^p & \tilde{\sigma}_{63}^p & \tilde{\sigma}_{64}^p & \tilde{\sigma}_{65}^p & \tilde{\sigma}_{66}^p \end{bmatrix} \quad (\text{A.34})$$

Note that in this fully flexible covariance matrix, none of the parameters have any economic content; in other words, none of the original covariance parameters can be recovered from estimating the parameters of this normalized covariance matrix. This is not a restriction and in fact allows for the most general substitution patterns possible.

Choice Probabilities

This subsection outlines how the choice probabilities are calculated if the individual chooses to search. The choice probabilities with no search are simpler, but less illustrative of the process, so will be provided later. The procedure used in this section is the GHK simulator (Geweke 1989; Hajivassiliou and McFadden 1998; Keane 1994).

It is easier to express the choice probabilities by introducing some matrix notation first. Utilities can be stacked as follows:

$$U_i = \begin{bmatrix} u_{i0}^s \\ u_{i1}^s \\ u_{i0}^p \\ u_{i1}^p \\ u_{i2}^p \\ u_{i3}^p \\ u_{i4}^p \\ u_{i5}^p \\ u_{i6}^p \end{bmatrix} \quad \tilde{U}_{i0} = \begin{bmatrix} u_{i1}^s - u_{i0}^s \\ u_{i1}^p - u_{i0}^p \\ u_{i2}^p - u_{i0}^p \\ u_{i3}^p - u_{i0}^p \\ u_{i4}^p - u_{i0}^p \\ u_{i5}^p - u_{i0}^p \\ u_{i6}^p - u_{i0}^p \end{bmatrix} = \begin{bmatrix} \tilde{u}_i^s \\ \tilde{u}_{i1}^p \\ \tilde{u}_{i2}^p \\ \tilde{u}_{i3}^p \\ \tilde{u}_{i4}^p \\ \tilde{u}_{i5}^p \\ \tilde{u}_{i6}^p \end{bmatrix} \quad (\text{A.35})$$

U_i denotes the vector of choice utilities and \tilde{U}_{i0} denotes the vector of differenced utilities where the second stage utilities are differenced against the outside option. More generally, let \tilde{U}_{ij} denote the vector of utilities where the second stage utilities are differenced against the j^{th} alternative.

The latent search and product choice utilities can be decomposed into observable and unobservable components. These are used to then rewrite \tilde{U}_{i0} in terms of observable and unobservable components:

$$u_{ik}^s = v_{ik}^s + \epsilon_{ik}^s \quad k \in \{0, 1\} \quad (\text{A.36})$$

$$u_{ij}^p = v_{ij}^p + \epsilon_{ij}^p \quad j \in \{0, 1, \dots, 6\} \quad (\text{A.37})$$

$$\tilde{U}_{i0} = \begin{bmatrix} \tilde{v}_i^s + \tilde{\epsilon}_i^s \\ \tilde{v}_{i1}^p + \tilde{\epsilon}_{i1}^s \\ \tilde{v}_{i2}^p + \tilde{\epsilon}_{i2}^s \\ \tilde{v}_{i3}^p + \tilde{\epsilon}_{i3}^s \\ \tilde{v}_{i4}^p + \tilde{\epsilon}_{i4}^s \\ \tilde{v}_{i5}^p + \tilde{\epsilon}_{i5}^s \\ \tilde{v}_{i6}^p + \tilde{\epsilon}_{i6}^s \end{bmatrix} \quad (\text{A.38})$$

Without loss of generality I will first explicitly write out the probability that an individual searches and chooses the outside option. This methodology will be used for each of the second stage alternatives in which an individual chooses to search (e.g., search/new/bookstore, search/used/bookstore, etc...).

$$Pr(u_{i1}^s < u_{i0}^s \ \& \ u_{i0}^p > u_{ik}^p \ \forall k \neq j \in \{1, \dots, 6\}) \quad (\text{A.39})$$

$$= Pr(\tilde{u}_i^s < 0 \ \& \ \tilde{u}_{ij}^p < 0 \ \forall j \in \{1, \dots, 6\}) \quad (\text{A.40})$$

$$= Pr(\tilde{\epsilon}_i^s < -\tilde{v}_i^s \ \& \ \tilde{\epsilon}_{i1}^p < -\tilde{v}_{i1}^p \ \& \ \tilde{\epsilon}_{i2}^p < -\tilde{v}_{i2}^p \ \& \ \dots \ \& \ \tilde{\epsilon}_{i6}^p < -\tilde{v}_{i6}^p) \quad (\text{A.41})$$

If the epsilons in Equation (A.41) were independent, then estimating this equation would be straightforward and not require simulation methods. However, recall that $\tilde{\epsilon}_i \sim \mathcal{N}(0, \tilde{\Sigma})$, where the epsilons are potentially correlated. In order to take draws from this distribution, the epsilons are transformed using the Choleski factor of $\tilde{\Sigma}$.

Let L_0 be the lower triangular matrix such that $\tilde{\Sigma}_0 = L_0 L_0'$.⁶

⁶In practice, c_{11} and c_{22} are normalized to one. A future section will describe how an initial Choleski factor is used to generate Σ from which each $\tilde{\Sigma}_j$ is derived. This initial Choleski factor contains the

$$L_0 = \begin{bmatrix} c_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\ c_{21} & c_{22} & 0 & 0 & 0 & 0 & 0 \\ c_{31} & c_{32} & c_{33} & 0 & 0 & 0 & 0 \\ c_{41} & c_{42} & c_{43} & c_{44} & 0 & 0 & 0 \\ c_{51} & c_{52} & c_{53} & c_{54} & c_{55} & 0 & 0 \\ c_{61} & c_{62} & c_{63} & c_{64} & c_{65} & c_{66} & 0 \\ c_{71} & c_{72} & c_{73} & c_{74} & c_{75} & c_{76} & c_{77} \end{bmatrix} \quad (\text{A.42})$$

Using the elements of L_0 , the elements of $\tilde{\epsilon}_{i0}$ can be rewritten as a linear combination of iid standard normal random variables, η_1, \dots, η_7 :

$$\tilde{\epsilon}_i^s = c_{11}\eta_1$$

$$\tilde{\epsilon}_{i1}^p = c_{21}\eta_1 + c_{22}\eta_2$$

$$\tilde{\epsilon}_{i2}^p = c_{31}\eta_1 + c_{32}\eta_2 + c_{33}\eta_3$$

$$\tilde{\epsilon}_{i3}^p = c_{41}\eta_1 + c_{42}\eta_2 + c_{43}\eta_3 + c_{44}\eta_4$$

$$\tilde{\epsilon}_{i4}^p = c_{51}\eta_1 + c_{52}\eta_2 + c_{53}\eta_3 + c_{54}\eta_4 + c_{55}\eta_5$$

$$\tilde{\epsilon}_{i5}^p = c_{61}\eta_1 + c_{62}\eta_2 + c_{63}\eta_3 + c_{64}\eta_4 + c_{65}\eta_5 + c_{66}\eta_6$$

$$\tilde{\epsilon}_{i6}^p = c_{71}\eta_1 + c_{72}\eta_2 + c_{73}\eta_3 + c_{74}\eta_4 + c_{75}\eta_5 + c_{76}\eta_6 + c_{77}\eta_7$$

Using this, the choice probability given in Equation (A.41) can be written in terms of the CDFs

of uncorrelated standard normals:

$$\begin{aligned}
& Pr(\tilde{\epsilon}_i^s < -\tilde{v}_i^s \text{ \& } \tilde{\epsilon}_{i1}^p < -\tilde{v}_{i1}^p \text{ \& } \tilde{\epsilon}_{i2}^p < -\tilde{v}_{i2}^p \text{ \& } \dots \text{ \& } \tilde{\epsilon}_{i6}^p < -\tilde{v}_{i6}^p) \\
& = Pr\left(\eta_1 < \frac{-\tilde{v}_i^s}{c_{11}}\right) \\
& \cdot Pr\left(\eta_2 < \frac{-\tilde{v}_{i1}^p - c_{21}\eta_1}{c_{22}} \middle| \eta_1 < \frac{-\tilde{v}_i^s}{c_{11}}\right) \\
& \cdot Pr\left(\eta_3 < \frac{-\tilde{v}_{i2}^p - c_{31}\eta_1 - c_{32}\eta_2}{c_{33}} \middle| \eta_1 < \frac{-\tilde{v}_i^s}{c_{11}}, \eta_2 < \frac{-\tilde{v}_{i1}^p - c_{21}\eta_1}{c_{22}}\right) \\
& \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\
& \cdot Pr\left(\eta_7 < \frac{-\tilde{v}_{i6}^p - c_{71}\eta_1 - c_{72}\eta_2 - c_{73}\eta_3 - c_{74}\eta_4 - c_{75}\eta_5 - c_{76}\eta_6}{c_{77}} \middle| \eta_1 < \frac{-\tilde{v}_i^s}{c_{11}}, \eta_2 < \dots\right)
\end{aligned} \tag{A.43}$$

$$\begin{aligned}
& = \Phi\left(\eta_1 < \frac{-\tilde{v}_i^s}{c_{11}}\right) \\
& \cdot \Phi\left(\eta_2 < \frac{-\tilde{v}_{i1}^p - c_{21}\eta_1}{c_{22}} \middle| \eta_1 < \frac{-\tilde{v}_i^s}{c_{11}}\right) \\
& \cdot \Phi\left(\eta_3 < \frac{-\tilde{v}_{i2}^p - c_{31}\eta_1 - c_{32}\eta_2}{c_{33}} \middle| \eta_1 < \frac{-\tilde{v}_i^s}{c_{11}}, \eta_2 < \frac{-\tilde{v}_{i1}^p - c_{21}\eta_1}{c_{22}}\right) \\
& \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\
& \cdot \Phi\left(\eta_7 < \frac{-\tilde{v}_{i6}^p - c_{71}\eta_1 - c_{72}\eta_2 - c_{73}\eta_3 - c_{74}\eta_4 - c_{75}\eta_5 - c_{76}\eta_6}{c_{77}} \middle| \eta_1 < \frac{-\tilde{v}_i^s}{c_{11}}, \eta_2 < \dots\right)
\end{aligned} \tag{A.44}$$

where Φ denotes the CDF of a standard normal random variable. The only remaining difficulty is simulating the conditional probabilities in Equation (A.45). The algorithm below provides the details for simulating this choice probability.

1. Calculate $\Phi^{-1}\left(\eta_1 < \frac{-\tilde{v}_i^s}{c_{11}}\right)$.
2. Take NS draws from a standard uniform random variable. Label each draw with superscript d :

$$\mu^d \sim U(0, 1) \tag{A.46}$$

3. Transform μ into a uniform distribution with upper-bound at $\Phi^{-1}\left(\eta_1 < \frac{-\tilde{v}_i^s}{c_{11}}\right)$:

$$\tilde{\mu}^d = \mu^d \cdot \Phi^{-1}\left(\eta_1 < \frac{-\tilde{v}_i^s}{c_{11}}\right) \forall d \quad (\text{A.47})$$

4. Calculate simulated values of η_1 for each d :

$$\eta_1^d = \Phi^{-1}(\tilde{\mu}^d) \quad (\text{A.48})$$

5. Repeat this process to take NS draws from η_2, \dots, η_6 .

6. For each d , use these values to approximate the choice probability:

$$\begin{aligned} P_{i0}^d &= \Phi\left(\eta_1 < \frac{-\tilde{v}_i^s}{c_{11}}\right) \\ &\cdot \Phi\left(\eta_2 < \frac{-\tilde{v}_{i1}^p - c_{21}\eta_1^d}{c_{22}}\right) \\ &\cdot \Phi\left(\eta_3 < \frac{-\tilde{v}_{i2}^p - c_{31}\eta_1^d - c_{32}\eta_2^d}{c_{33}}\right) \\ &\quad \vdots \quad \quad \quad \vdots \\ &\cdot \Phi\left(\eta_7 < \frac{-\tilde{v}_{i6}^p - c_{71}\eta_1^d - c_{72}\eta_2^d - c_{73}\eta_3^d - c_{74}\eta_4^d - c_{75}\eta_5^d - c_{76}\eta_6^d}{c_{77}}\right) \end{aligned} \quad (\text{A.49})$$

7. The simulated choice probability is given by,

$$\frac{1}{NS} \sum_{d=1}^{NS} P_{i0}^d \quad (\text{A.50})$$

The steps laid out in this section are used to calculate each choice probability for which an individual chooses to search first. Calculating choice probabilities for the case in which an individual does not search is even simpler; in this case, the probability of each of the online alternatives is zero, so this reduces the number of simulated conditional probabilities shown in Equation (A.49)

by three⁷.

Likelihood Function

Using the simulated choice probabilities given in the previous subsection, the parameters governing the search decision and the parameters of the alternative specific utility function, collectively Θ , can be estimated using simulated maximum likelihood estimation. The individual likelihood function is given by,

$$\mathbb{L}_i(\Theta) = \prod_j P_{ij}^{d_{ij}} \quad (\text{A.51})$$

for $j \in \{1, \dots, 11\}$. The simulated choice probabilities are given by P_{ij} and d_{ij} is an indicator function equal to one if individual i chooses choice path j ⁸.

The full log-likelihood function used in estimation is given by,

$$\mathcal{L}(\Theta) = \sum_i \sum_j d_{ij} \ln(\tilde{P}_{ij}) \quad (\text{A.52})$$

A.2 Data

A.2.1 Price Issues

Students were asked to report the price they paid for their textbook as part of the online questionnaire. Because this is sure to be measured with error and bookstore prices are observable and static over the course of the semester, the magnitude of this measurement error can be quantified for the students who purchased books at the campus bookstore. Assuming this measurement error is similar for students who purchase textbooks online, then the estimate of the bias and variance of the measurement error can be used to correct for students' reported price for online retailers. The static nature of the bookstore prices also allow me to impute accurate prices for these options.

⁷In the future I intend to further elaborate on how Σ is initially calculated, how exogenous variation in new or used bookstore alternatives affects the choice probabilities, and how matrix transformations are used to calculate \tilde{U} and $\tilde{\Sigma}$.

⁸Recall that a choice path is a combination of two decisions: a search decision and a product choice. There are 7 potential choices if an individual chooses to search and 4 if an individual does not search.

For bookstore purchases, prices collected directly from those retailers and not the price reported by students on the online questionnaire will be used in estimation. For students who purchase or rent textbooks from online retailers, this is more difficult. Prices for textbooks in online marketplaces vary widely over the course of a semester. Since accurate price information is vital to this study's results, price and quality information for each assigned textbook was collected daily from eleven online retailers, two online publishers, and the website of the other college bookstore in the Chapel Hill area⁹. One of two strategies will be used to assign a price of an online book to a student.

The first, and preferable case, is that students do not report textbook prices with a significant amount of error (it is worth reiterating that this can be quantified by comparing actual bookstore prices with the price reported by students who purchased from a bookstore). In this case, a distributional assumption will be made on measurement error and this will be accounted for in estimation.

The second case is that students report textbook prices with significant amounts of error: this could be a significant bias and/or large amounts of variability. In this case, data scraped from online retailers will be used to impute a price. For example, if a student reports having purchased a used copy of Mankiw's Principles of Economics textbook with "very good" quality from Amazon.com two days before class started, then the lowest price of a used-very good textbook on Amazon.com for that day will be imputed as being the purchase price.

A.2.2 Online Questionnaire

The online questionnaire contains 68 unique questions. Many of them are conditional on past responses, so the actual number of questions students were asked to respond to varied. Listed below are the 68 items, their potential responses, and the conditions on the questions being prompted (conditions given in italics).

1. *All*: Which of the following choices best summarizes your decision for this course, (Course

⁹The online retailers include online marketplaces to purchase textbooks (Amazon.com, Half.com, Barnes and Noble, Abebooks, textbooks.com, textbooksrus.com), retailers specializing in textbook rentals (bookrenter.com, campusbookrentals.com, campusbuddy.com, chegg.com, and collegebookrenter.com), online publishers (lulu.com and universityreaders.com) and Ram Book and Supply's website.

Name) regarding the assigned book: (Book Title) by (Author) ?

- (a) Purchased the book
- (b) Rented the book
- (c) Borrowed the book from another person
- (d) Use a copy from the library
- (e) Do not have the book
- (f) Other

1.1 *If “Purchased the book”*: Where did you purchase the book?

- (a) UNC Student Stores (campus bookstore)
- (b) Ram Book & Supply
- (c) An online retailer (ex. Amazon.com)
- (d) A friend or classmate
- (e) Other

1.1.1 *If “An online retailer”*: Please write in the name of the online retailer.

(a) _____

1.1.2 *If “Other”*: Please write in where you purchased the book:

(a) _____

1.2 *If “Purchased the the book”*: When did you purchase the book? (ex. four days before class started)

(a) _____

1.3 *If “Purchased the the book”*: How much did the book cost (including tax and/or shipping)?

(a) _____

1.4 *If “Rented the book”*: Where did you rent the book?

- (a) Ram Book & Supply
- (b) An online retailer (ex. Amazon.com)
- (c) Other

1.4.1 *If “An online retailer”*: Please write in the name of the online retailer.

(a) _____

1.4.2 *If “Other”*: Please write in where you rented the book:

(a) _____

1.5 *If “Rented the book”*: How much did the book cost (including tax and/or shipping)?

(a) _____

1.6 *If “Rented the book”*: How many days did you rent the book for?

(a) _____

1.7 *If “Rented the book”*: When did you rent the book? (ex. four days before class started)

(a) _____

1.8 *If “Use a copy from the library”*: Did you check out the book from the library or do you use a reserve copy?

(a) Checked out

(b) Reserve copy

1.9 *If “Other”*: Please provide a short description of how you obtained the book for this course:

(a) _____

1.10 *If “Do not have the book”*: Do you plan on getting the book?

(a) Yes

(b) No

1.10.1 *If “Yes”* How do you plan on getting the book?

- (a) Purchasing the book
- (b) Renting the book
- (c) Borrowing the book from a friend
- (d) Checking the book out from a library
- (e) Other

2. *If “Purchased or Rented”*: Do you share the book with another student?

- (a) Yes
- (b) No

3. *If “Purchased from UNC Student Stores” is not selected*: Did you check the price of the book at the UNC Student Stores (campus bookstore)?

- (a) Yes
- (b) No

4. *If “Purchased or rented from Ram Book & Supply” is not selected*: Did you check the price of the book at a bookstore other than the UNC Student Stores (campus bookstore)?

- (a) Yes, I priced this book at another bookstore
- (b) No
- (c) Yes, but the book is not available at other bookstores

4.1 *If “Did not check the price at UNC Student Stores or another bookstore” are both selected*: How much do you think a new copy of the book costs at the UNC Student Stores?

- (a) _____

4.2 *If “Did not check the price at UNC Student Stores or another bookstore” are both selected*: How much do you think a used copy of the book costs at the UNC Student Stores?

(a) _____

5. *If “Purchased or rented from an online retailer” is not selected:* Did you check the price of the book at an online retailer (ex. Amazon.com)?

(a) Yes, I priced this book at an online retailer

(b) No

(c) Yes, but the book is not available at online retailers

5.1 *If “Did not check the price at an online retailer” is selected:* If you had looked at online retailers, what do you think is the lowest price you would be able to find to buy a new copy of this book?

(a) _____

5.2 *If “Did not check the price at an online retailer” is selected:* If you had looked at online retailers, what do you think is the lowest price you would be able to find to buy a used copy of this book?

(a) _____

5.3 *If “Did not check the price at an online retailer” and “Rented the book” are selected:* If you had looked at online retailers, what do you think is the lowest price you would be able to find to rent a copy of this book?

(a) _____

6. *All:* How did you determine what book was assigned for this course?

(a) UNC Student Stores website

(b) Visited UNC Student Stores

(c) Another website

(d) Through the professor

(e) Other

6.1 *If “Other”*: Please briefly describe how you determined what book was assigned for this course.

(a) _____

7. *If “Rented the book” is not selected*: At the beginning of the semester, did you know that certain bookstores and online retailers offer textbook rentals?

(a) Yes

(b) No

8. *If “Purchased the book”*: Even if you plan on keeping your book at the end of the semester, what do you think the resale value of your book will be at the end of the semester?

(a) \$ _____

9. *If “Purchased the book”*: Even if you plan on keeping your book at the end of the semester, what is the lowest amount you would be willing to sell your book for, once you are finished taking this course? (Note that this is different than the previous question)

(a) \$ _____

10. *If “Purchased the book”*: Which of the following do you think you are most likely to do with your book at the end of the semester?

(a) Sell to a bookstore

(b) Sell online

(c) Keep the book for future use

(d) Other

10.1 *If “Other”*: Please write a short description of what you intend to do with your book after this semester is over.

(a) _____

11. *If “Purchased the book”*: Do you think the bookstore will be buying back the book for this course at the end of the semester?

(a) Definitely yes

(b) Maybe

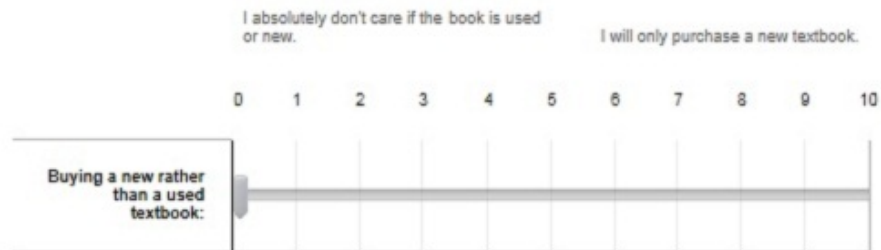
(c) Definitely not

11.1 *If “Maybe”*: What do you think is the likelihood that the bookstore will buy the book back? (ex. I think there is a _____% chance they will buy back the book)

(a) I think there is a _____% chance the bookstore will buy back the book.

The following page contains images from the actual survey. Students were prompted to use a sliding scale to answer 5 questions regarding textbook preferences.

On a scale of 0 to 10, how important are the following to you?



On a scale of 0 to 10, how confident are you that you will be satisfied with an online purchase?



12. If purchased the book from UNC Student Stores or Ram Book & Supply: Did you buy a new

or used book?

- (a) New
- (b) Used

12.1 *If “Used”*: Please choose the response that fits the condition of your used book best (when you bought it).

- (a) Like New
- (b) Very Good
- (c) Good
- (d) Acceptable
- (e) Poor

12.2 *If “Used”*: Was a new option also available?

- (a) Yes
- (b) No

12.3 *If “New”*: Was a used option also available?

- (a) Yes
- (b) No

13. *If purchased the book from an online retailer*: What was the advertised condition of the book when you bought it? Please choose the response that fits your textbook best.

- (a) New
- (b) Used - Like New
- (c) Used - Very Good
- (d) Used - Good
- (e) Used - Acceptable

14. *If purchased the book:* Is the book you bought the edition assigned to the course or a previous edition?
- (a) Current Edition
 - (b) Previous Edition
15. *If rented the book:* Is the book you rented the edition assigned to the course or a previous edition?
- (a) Current Edition
 - (b) Previous Edition
16. *If purchased or rented the book:* Which of the following options best describes your book?
- (a) Hardcover
 - (b) Paperback
 - (c) Loose-leaf pages
 - (d) E-book
 - (e) International Edition
17. *If purchased or rented the book from an online retailer:* Did you choose standard or expedited shipping?
- (a) Standard
 - (b) Expedited
18. *If purchased or rented the book:* Did your book come with bundled materials (ex. CDs, study guides, online access codes, etc.)?
- (a) Yes
 - (b) No
19. *If purchased or rented the book:* Who paid for your book?

- (a) Self
- (b) Another Person (ex. Parent)
- (c) Scholarship
- (d) Other

19.1 *If “Another Person”*: Did you decide where to get the book or did the other person?

- (a) I chose
- (b) The other person chose

19.2 *If “Another Person”*: Did the other person give you a set amount to spend on books this semester (ex. \$400 towards textbooks) or did the other person pay a varying amount, depending on the price of the book (ex. a check or credit card to pay for the entire cost of books)?

- (a) A set amount
- (b) A varying amount

19.3 *If “Scholarship” and purchased from the UNC Student Stores*: Does your scholarship allow you to get books from the UNC Student Stores without physically paying for them that day?

- (a) Yes
- (b) No

19.4 *If “Other”*: Please write a short description of who paid for your book.

- (a) _____

20. *All*: Counting a summer session or maymester as one semester, how many semesters, including this one, have you attended UNC and/or another college?

- (a) _____ semesters

21. *All*: According to the University, which of the following are you currently classified as?

- (a) Freshman
- (b) Sophomore
- (c) Junior
- (d) Senior
- (e) Graduate Student
- (f) Other
- (g) Don't Know

21.1 *If "Other"*: How does the University classify you, if not as a freshman, sophomore, etc.?

(a) _____

22. *All*: At the beginning of the semester, were you classified as an in-state or out-of-state student?

- (a) In-state
- (b) Out-of-state

23. *All*: Do you live on-campus or off-campus?

- (a) On-campus
- (b) Off-campus

24. *All*: Are you male or female?

- (a) Male
- (b) Female

25. *All*: How old are you?

(a) _____years

26. *All*: If you have siblings, are you the first to attend college?

- (a) Yes
- (b) No
- (c) Not Applicable

27. *All:* Is this a course in your major field?

- (a) Yes
- (b) No
- (c) Undeclared

28. *All:* Counting a summer session or maymester as one semester, how many semesters, including this one, have you attended UNC and/or another college?

- (a) _____ semesters

29. *All:* Other than this semester, have you ever purchased a college textbook?

- (a) Yes
- (b) No

30. *All:* Have you ever purchased anything online?

- (a) Yes
- (b) No

30.1 *If "Yes":* Have you ever had a negative online shopping experience?

- (a) Yes
- (b) No

31. *All:* Have you ever sold anything online?

- (a) Yes
- (b) No

32. *All:* Would you like to be informed of future opportunities to participate in short online surveys for money and/or prizes?

(a) Yes

(b) No

33. *All:* Please provide your email address to be entered for a chance to win one of five \$100 cash prizes

(a) _____

APPENDIX A

APPENDIX TO CHAPTER 3: PRICE BELIEFS AND EXPERIENCE

A.1 Data Appendix

A.1.1 Online Questionnaire Data

Individuals who participated in the online questionnaire were respondents from a list of emails generated through participation in a separate online questionnaire conducted at UNC during the 2011-2012 and 2012-2013 academic calendar years. These individuals agreed to participate in follow-up surveys at the completion of the separate questionnaire. This separate questionnaire was distributed by instructors to their students, who had the option to participate. Additional details about this separate questionnaire are available in Spence (2013).

Online Questionnaire

[The following is a subset of the questions provided to textbook consumers using Qualtrics online survey software. Notes are provided in brackets.]

Textbook Purchasing Questionnaire

The following survey seeks to gain understanding into how consumers choose which retailers to consider when faced with purchasing decisions. Over the course of this survey you will be presented with a number of hypothetical textbook purchasing decisions. You will be asked about your price expectations from online retailers and your beliefs about the time costs involved with searching within an online market. You will not actually have to price any textbooks from online retailers or visit any website outside of this survey.

Directions: Please answer all questions to the best of your ability. Use the right arrow button at the bottom of the screen to advance to the next page. You may also use the left arrow at the bottom of the screen to move back at any time and change a previous answer. If you are uncomfortable answering a specific question you can either skip that question or exit the survey. Thank you for participating!

How many semesters in total, including this one, have you attended UNC and any other college? (Count a summer session as a semester)

Semesters: _____

What is (are) your major(s)? Please write Undecided if you do not currently have a stated major.

Major(s): _____

Which of the following have you ever purchased a textbook from? (Please check all that apply)

- ☐ UNC Student Stores (campus bookstore)
- ☐ Ram Book and Supply
- ☐ Another college bookstore
- ☐ Amazon.com
- ☐ Half.com or Ebay.com
- ☐ Ecampus.com
- ☐ Chegg.com
- ☐ Another online retailer
- ☐ Another student (directly)

Which of the following have you ever rented a textbook from? (Please check all that apply)

- ☐ I have never rented a textbook
- ☐ UNC Student Stores (campus bookstore)
- ☐ Ram Book and Supply
- ☐ Another college bookstore
- ☐ Amazon.com
- ☐ Half.com or Ebay.com
- ☐ Ecampus.com
- ☐ Chegg.com
- ☐ Another online retailer
- ☐ Another student (directly)

Please write in the other online retailers you have ever rented or purchased a textbook from:

When do you normally purchase (or order) your textbooks?

- ☐ More than 2 weeks before the semester starts
- ☐ 1 - 2 weeks before the semester starts

- ☐ A few days before the semester starts
- ☐ The day the semester starts
- ☐ A few days after the semester starts
- ☐ 1 - 2 weeks after the semester starts
- ☐ More than 2 weeks after the semester starts

Have you purchased or rented any textbooks for an upcoming summer session?

- ☐ Yes
- ☐ No

How many textbooks have you ever purchased or rented online?

- ☐ 1 - 5
- ☐ 6 - 10
- ☐ 11 or more

When purchasing or renting a textbook online, have you ever used a website that shows the lowest prices available from multiple online retailers?

- ☐ Yes
- ☐ No

On average, when you purchase a textbook online, how many different online retailers do you visit? Number of Retailers Visited: _____

Do you have an Amazon Prime membership?

- ☐ Yes
- ☐ No

Do you have a Paypal account?

- ☐ Yes
- ☐ No

How many online purchases do you typically make in a three month period?

Number of Purchases: _____

If you were given an isbn number or textbook title and wanted to purchase or rent this textbook online, what

is the first website you would visit?

Website Name: _____

Hypothetically, if you only visited one online retailer, how many minutes do you think it would take to look up **one** textbook and purchase it? (Include the time to search, find the option you want, enter your information, and complete the transaction)

Minutes: _____

Hypothetically, if you only visited one online retailer, how many minutes do you think it would take to look up **three** textbooks and purchase them? (Include the time to search, find the option you want, enter your information, and complete the transaction)

Minutes: _____

You will now be given a number of hypothetical textbook purchasing decisions. In each case, you will be given information about the textbook and asked to give your best guess about what the price of this textbook is from an online retailer. This survey is concerned about what your expectations are about prices from online retailers, so please do not actually search for the price of the textbook online. Before presenting you with the hypothetical purchasing decisions, you will be provided with an example of what the questions will be like.

Example: If you searched only one clothing store in the mall (ex. Old Navy), what do you think is the lowest price you could find for a pair of jeans in your size? Please enter your answer as a number. Note that this question does not have a right or wrong answer, it simply asks for your best guess.

\$_____ [Denoted “Example Expectation” in following questions]

Example Continued: Given that you don't know the lowest price of a pair of jeans with certainty, there is some chance that the lowest price is lower than \$[Example Expectation] and some chance that the lowest price is greater than \$[Example Expectation]. In the following questions, you will be asked about your beliefs about the chance that the lowest price you could find would be below \$[Example Expectation] and also the chance the lowest price you could find would be above \$[Example Expectation].

What do you think is the chance that the lowest price of the pair of jeans is less than \$[90% of Example Expectation]? Please enter the chance as a percentage (i.e. a number between 0 and 100). For example:

I think there is a 30% chance that the lowest price of the pair of jeans is less than \$[90% of Example Expectation];

Percent Chance: _____

What do you think is the chance that the lowest price of the pair of jeans is more than \$[110% of Example Expectation]? Please enter the chance as a percentage (i.e. a number between 0 and 100). For example: “I think there is a 35% chance that the lowest price of the pair of jeans is greater than \$[Example Expectation].”

Percent Chance: _____

You will now be given three textbook purchasing scenarios, each similar to the previous example.

[The following is an example using one of the possible twelve textbooks. Respondents were given three scenarios randomly drawn from three groups of four textbooks (one from each group).]

Scenario: You are assigned “Economics: Principles and Policy” by William Baumol and Alan Blinder for an introductory economics course (ECON-101). [The following information on textbook characteristics was randomly assigned to respondents with 50% likelihood. The likelihood of receiving this information for the subsequent scenarios did not depend on whether the information on characteristics was shown for previous textbook purchasing scenarios.] This is the twelfth and latest edition of the textbook, it was published by South-Western College Publishing, and was last revised in 2012. The dimensions of the book are 8.4 x 1.5 x 11.1 inches, it is a hardcover, it contains 880 pages, and it weighs 4.4 pounds. A picture is provided below:

[Picture presented such as the one shown in the screenshot in Section 3]

Have you ever taken this course?

☐ Yes

☐ No

Have you ever been assigned this textbook?

☐ Yes

☐ No

You know that a new copy of this book costs \$212 (including taxes) at the UNC Student Stores. If you searched one online retailer, what do you think the price of a new copy at this online retailer would be (include shipping costs)? Reminder: Please do not actually search for this price. Provide your best guess

instead.

\$ _____ [Denoted “New Expectation” in future questions]

What do you think is the probability that the lowest price for a new copy of this book is less than \$[85%, 90%, or 95% of New Expectation] (including shipping costs) at the online retailer?

Percent Chance: _____

What do you think is the probability that the lowest price for a new copy of this book costs more than \$[105%, 110%, or 115% of New Expectation] (including shipping costs) at the online retailer? Note that your answer to this question added to your answer from the previous question should not exceed 100.

Percent Chance: _____

You know that a used copy of this book costs \$159 (including taxes) at the UNC Student Stores. If you searched one online retailer, what do you think the price of a used copy at this online retailer would be (include shipping costs)?

\$ _____ [Denoted “Used Expectation” in future questions]

What do you think is the probability that the lowest price for a used copy of this book costs less than \$[85%, 90%, or 95% of Used Expectation] (including shipping costs) at the online retailer?

Percent Chance: _____

What do you think is the probability that the lowest price for a used copy of this book costs more than \$[105%, 110%, or 115% of Used Expectation] (including shipping costs) at the online retailer? Note that your answer to this question added to your answer from the previous question should not exceed 100.

Percent Chance: _____

A.1.2 Textbook Purchasing Scenarios

Table A.1 provides information on the textbooks used in the hypothetical textbook purchasing scenarios. Respondents that completed the survey faced three scenarios; in each scenario, one textbook from each group was randomly assigned to the respondent. The first group is composed of social science textbooks; the second group is composed of hard science textbooks; the third group is composed of humanities textbooks. In the Fall 2012 semester, individuals were presented these scenarios in the previous ordering (social sciences, hard sciences, then humanities). The Spring

Table A.1: : Textbook Scenarios

| Book # | Group | Title | | Author | Edition | Latest Edition |
|--------|-------|--|---------------------------|--------|---------|----------------|
| 1 | 1 | Economics: Principles and Policies | Baumol and Blinder | | 12 | Yes |
| 2 | 1 | Introductory Econometrics: A Modern Approach | Wooldridge | | 4 | No |
| 3 | 1 | Experience Sociology | Croteau and Hoynes | | 1 | Yes |
| 4 | 1 | Medical Sociology | Cockerham | | 12 | Yes |
| 5 | 2 | Chemistry: The Essential Science | Brown et al. | | 12 | Yes |
| 6 | 2 | Animal Physiology: Adaptation and Environment | Schmidt-Nielsen | | 5 | Yes |
| 7 | 2 | Earth: Portrait of a Planet | Marshak | | 3 | No |
| 8 | 2 | Data Structures and Algorithm Analysis in Java | Weiss | | 3 | Yes |
| 9 | 3 | Norton Anthology of Short Fiction | Bausch and Cassill | | 7 | Yes |
| 10 | 3 | Medicine and Morality in Haiti | Brodwin | | 1 | Yes |
| 11 | 3 | Western Civilization, Volume 1 | Perry | | 10 | Yes |
| 12 | 3 | Voces de Hispanoamerica | Chang-Rodrigues and Filer | | 4 | Yes |

| Book # | Year Published | Subject | Course | Cover | Pages | Weight (lbs) | ISBN | New Bookstore Price |
|--------|----------------|------------------|--------|-------|-------|--------------|---------------|---------------------|
| 1 | 2012 | Economics | 101 | Hard | 880 | 4.4 | 9780538453577 | \$286 |
| 2 | 2008 | Economics | 570 | Hard | 896 | 3.4 | 9780324581621 | \$262 |
| 3 | 2012 | Sociology | 101 | Paper | 576 | 2.8 | 978007319353 | \$141 |
| 4 | 2011 | Sociology | 469 | Paper | 432 | 1.2 | 9780205054183 | \$87 |
| 5 | 2011 | Chemistry | 101 | Hard | 1200 | 5.4 | 9780321696724 | \$253 |
| 6 | 1997 | Biology | 451 | Hard | 617 | 3.4 | 9780521570985 | \$104 |
| 7 | 2007 | Geology | 101 | Paper | 880 | 4.4 | 9780393935189 | \$146 |
| 8 | 2011 | Computer Science | 410 | Hard | 640 | 2.1 | 9780132576277 | \$151 |
| 9 | 2006 | English | 123 | Paper | 1776 | 2.8 | 9780393926118 | \$81 |
| 10 | 1996 | Anthropology | 470 | Paper | 260 | 1.0 | 9780521575430 | \$52 |
| 11 | 2012 | History | 151 | Paper | 496 | 1.6 | 9781111831707 | \$172 |
| 12 | 2012 | Spanish | 400 | Hard | 706 | 2.2 | 9781111837921 | \$213 |

2013 questionnaire assigned individuals to groups at random (i.e. roughly one third of respondents completed a scenario with a hard sciences textbook, then social sciences, then humanities).

Textbooks were chosen to provide variation in the following characteristics: the number of total editions of the textbook, whether the textbook is the latest edition, the year of publication, whether the course was designed for an introductory or upper-level course, the type of cover (hardback vs. paperback), the number of pages, and the weight. In the following tables, course number refers to the numbering at UNC for the Fall 2012 and Spring 2013 semesters. New bookstore price refers to the price from UNC's campus bookstore during these semesters.

A.1.3 Survey Attrition and Estimation Sample

Table A.2 summarizes the number of respondents at various points within the survey. For the Fall 2012 semester, 979 respondents began the questionnaire and 734 (75%) completed the questionnaire. For the Spring 2013 semester, 1002 respondents began the questionnaire and 703 (70%) completed the questionnaire. We only exclude individuals who did not complete the background questions. This leaves 820 respondents from the Fall and 798 respondents from the Spring. Of these individuals, we exclude 104 respondents (52 from both semesters) because they had been enrolled in college for more than 10 semesters and/or summer sessions and an additional 49 respondents (29 from Fall 2012 and 20 from Spring 2013) for reporting nonsensical answers (e.g., reporting an expected price of \$100,000).

Table A.2: : Survey Attrition

| | Fall 2012 Respondents | Percent Remaining | Spring 2013 Respondents | Percent Remaining |
|------------------------------------|--------------------------|----------------------|----------------------------|----------------------|
| Began the questionnaire | 979 | 100 | 1002 | 100 |
| Completed the background questions | 820 | 83.8 | 798 | 79.6 |
| Completed at least one scenario | 759 | 77.5 | 761 | 75.9 |
| Completed at least two scenarios | 741 | 75.7 | 716 | 71.5 |
| Completed the questionnaire | 734 | 75.0 | 703 | 70.2 |

A.2 Robustness Checks

This section investigates the robustness of the results presented in the paper by providing results from a number of other specifications. Explicitly, we explore the robustness of our results by

Table A.3: : Results Using Alternative Samples

| Experience | Used Books | | | | | | | | |
|-----------------------|-------------|-------|-------|-----------------|-------|-------|-----------------|-------|-------|
| | Main Sample | | | Extended Sample | | | Single Scenario | | |
| | N | Mean | SD | N | Mean | SD | N | Mean | SD |
| No Online Purchases | 438 | 0.739 | 0.201 | 450 | 0.732 | 0.211 | 151 | 0.736 | 0.209 |
| 1-5 Online Purchases | 917 | 0.712 | 0.181 | 939 | 0.709 | 0.189 | 311 | 0.700 | 0.190 |
| 6-10 Online Purchases | 957 | 0.683 | 0.173 | 976 | 0.677 | 0.181 | 324 | 0.674 | 0.181 |
| 11+ Online Purchases | 1698 | 0.653 | 0.180 | 1755 | 0.647 | 0.191 | 577 | 0.638 | 0.186 |
| Experience | New Books | | | | | | | | |
| | Main Sample | | | Extended Sample | | | Single Scenario | | |
| | N | Mean | SD | N | Mean | SD | N | Mean | SD |
| No Online Purchases | 440 | 0.833 | 0.172 | 452 | 0.832 | 0.186 | 152 | 0.828 | 0.177 |
| 1-5 Online Purchases | 923 | 0.819 | 0.153 | 945 | 0.814 | 0.161 | 313 | 0.804 | 0.166 |
| 6-10 Online Purchases | 966 | 0.783 | 0.161 | 985 | 0.779 | 0.165 | 326 | 0.761 | 0.176 |
| 11+ Online Purchases | 1706 | 0.770 | 0.157 | 1763 | 0.768 | 0.166 | 581 | 0.752 | 0.168 |

varying the following:

A.2.1 The criteria for begin omitted from the sample and the number of scenarios used for each respondent.

A.2.2 Testing the distributional assumption.

A.2.1 Omission Criteria

Respondents are omitted from our sample in the main body of the paper for two reasons:

1. Being enrolled in more than 10 semesters of college.
2. Reporting expectations less than 10% of the bookstore price or greater than 150% of the bookstore price.

The first criteria is used to focus on traditional college students. The second criteria is used to eliminate respondents who we believe did not take the questionnaire seriously (for example, individuals who reported expectations of \$0 or \$100,000). To make sure that our results are not

biased because of these omissions, we relax the second omission criteria. We also report evidence that our omission criteria is not correlated with our measures of experience.

We proceed to report the main findings from the paper for a less stringent omission criteria. Specifically, we only omit respondents who report expectations less than 1% of the bookstore price or greater than 200% of the bookstore price. This results in four respondents being omitted from the Fall sample and one respondent being omitted from the Spring sample for reporting expectations below 1% of the bookstore price, and four respondents being omitted from the Fall sample and four respondents being omitted from the Spring sample for reporting expectations greater than 200% of the bookstore price.

Table A.3 reports the mean ratio of expectations to bookstore prices for the main sample as well as the extended sample. Including outliers does not significantly change the estimates of mean price expectations. Also included in table A.3 are the results that only use the first hypothetical textbook purchasing scenario that an individual responded to (out of a potential of six for individuals who completed the questionnaire in the fall and spring semester). The price expectations for the first scenario are lower than for the full sample, but the relationship between experience and price expectations is the same for both groups.

Results from regressions of the ratio of expectations to bookstore prices on measures of experience and other covariates also remain quantitatively similar to the results reported in the main body of the paper (not reported).

Table A.4 reports the results for the normal and gamma distributions. Also included are the results for the log-normal and gamma distributions using a more restrictive sample. Since the normal distribution requires dropping individuals who report a greater than 50% probability of being below the threshold, the restricted samples are constructed using a similar rejection criteria. The samples are constructed by dropping individuals who report a 50% or greater probability of being below the threshold. Then the parameter values are calculated for each individual, and the final sample includes individuals whose parameter values are not in the top or bottom 2.5% of values for either parameter. The results for the log-normal and gamma using the restricted

Table A.4: : Distribution Results

| Experience | Used | | | New | | |
|--------------------------------|------|------------------------|---------------------------------|------|------------------------|---------------------------------|
| | N | mean $\mathbf{E}_i(p)$ | mean $\sqrt{\mathbf{Var}_i(p)}$ | N | mean $\mathbf{E}_i(p)$ | mean $\sqrt{\mathbf{Var}_i(p)}$ |
| Normal | | | | | | |
| No online purchases | 291 | 0.720*** (0.159) | 0.126 (0.096) | 300 | 0.837*** (0.134) | 0.139 (0.099) |
| 1-5 online purchases | 619 | 0.714*** (0.155) | 0.130* (0.094) | 662 | 0.822*** (0.129**) | 0.141* (0.096) |
| 6-10 online purchases | 667 | 0.701* (0.150***) | 0.125 (0.095) | 696 | 0.796 (0.130) | 0.130 (0.088*) |
| 11+ online purchases | 1202 | 0.669*** (0.160***) | 0.119** (0.089) | 1259 | 0.784*** (0.134*) | 0.133 (0.096) |
| Gamma | | | | | | |
| No online purchases | 411 | 0.747*** (0.192) | 0.219 (0.281) | 412 | 0.836*** (0.164) | 0.237 (0.288) |
| 1-5 online purchases | 852 | 0.717*** (0.175) | 0.244** (0.294**) | 870 | 0.819*** (0.152*) | 0.243* (0.312) |
| 6-10 online purchases | 900 | 0.689 (0.170***) | 0.234 (0.292) | 908 | 0.785* (0.158) | 0.224 (0.276) |
| 11+ online purchases | 1601 | 0.656*** (0.174) | 0.207*** (0.252***) | 1604 | 0.774*** (0.153) | 0.217* (0.269) |
| Log-Normal - Restricted Sample | | | | | | |
| No online purchases | 283 | 0.726*** (0.158) | 0.113 (0.068) | 299 | 0.836*** (0.135) | 0.129 (0.074***) |
| 1-5 online purchases | 610 | 0.715*** (0.156) | 0.116*** (0.067**) | 649 | 0.822*** (0.126***) | 0.131*** (0.071) |
| 6-10 online purchases | 647 | 0.701* (0.151***) | 0.110 (0.063) | 678 | 0.797 (0.128) | 0.121 (0.065**) |
| 11+ online purchases | 1168 | 0.665*** (0.161***) | 0.104*** (0.060**) | 1218 | 0.783*** (0.134**) | 0.121* (0.068) |
| Gamma - Restricted Sample | | | | | | |
| No online purchases | 317 | 0.754*** (0.197) | 0.108 (0.060) | 325 | 0.847*** (0.164) | 0.126** (0.072***) |
| 1-5 online purchases | 637 | 0.715*** (0.174) | 0.112*** (0.064**) | 690 | 0.821*** (0.151*) | 0.125*** (0.070**) |
| 6-10 online purchases | 676 | 0.694 (0.166***) | 0.107 (0.063) | 708 | 0.788** (0.153) | 0.114* (0.063**) |
| 11+ online purchases | 1210 | 0.660*** (0.174) | 0.101*** (0.059**) | 1263 | 0.780*** (0.150) | 0.114*** (0.063**) |

Notes: Standard deviations in parenthesis.

The significance levels reported for the mean values are from a two-sample equality of means test. The significance levels for the standard deviations are from Brown and Forsythe's alternative formulation of Levene's robust two-sample equality of variances test.

* refers to p-value < .1; ** < .05; *** < .01

samples does not change the mean of the price expectations by a large amount, but the variability of the expected price distribution falls substantially. With a similar sample construction, the normal distribution is closest to the variability of the empirical distribution. The higher mean variability of the expected price distribution using the less restrictive sample for the log-normal and gamma distributions is driven by the individuals who report a high probability of the price being below the threshold. Ultimately, the results are similar regardless of the distribution used.

A.2.2 Price Distribution

In this section we provide some evidence supporting the use of the log normal price distribution as well as discussing some limitations of the distribution in fitting certain features of the empirical price distribution. Tests for normality reject the assumption of normality for both the distribution of prices and the log of prices for most specifications of the empirical distribution. For used books with a list price greater than \$100, the assumption of normality cannot be rejected. For new books, the normal distribution is able to fit the data better than the log-normal distribution. For used books, the normal distribution only fits better for the relatively expensive books. In order for the log normal distribution to fit the long right tail of the price distribution, the result is that it places too little weight on the left tail relative to the empirical distribution. The analysis in this paper is not dependent on a particular distributional assumption. In structural search models, however, an incorrect distributional assumption on the individual's beliefs about the price distribution or about the empirical price distribution can significantly bias estimates. Figure A.1 displays histograms of the empirical prices for the Fall survey period. Figures A.2 and A.3 display kernel density estimates for used and new prices. Finally figures A.5 and A.4 display the time series of the mean daily price with the 95% confidence interval. During the survey periods, prices are relatively stable. Used prices rise considerably leading up to the start of the semester, however, new prices remain fairly stable throughout the year.

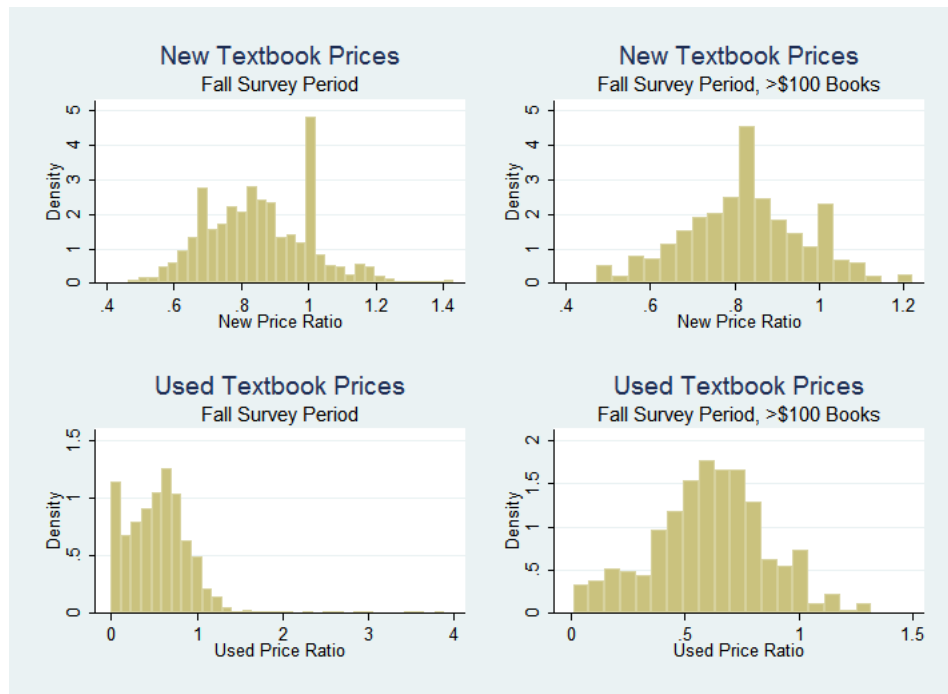


Figure A.1: Histograms of Prices

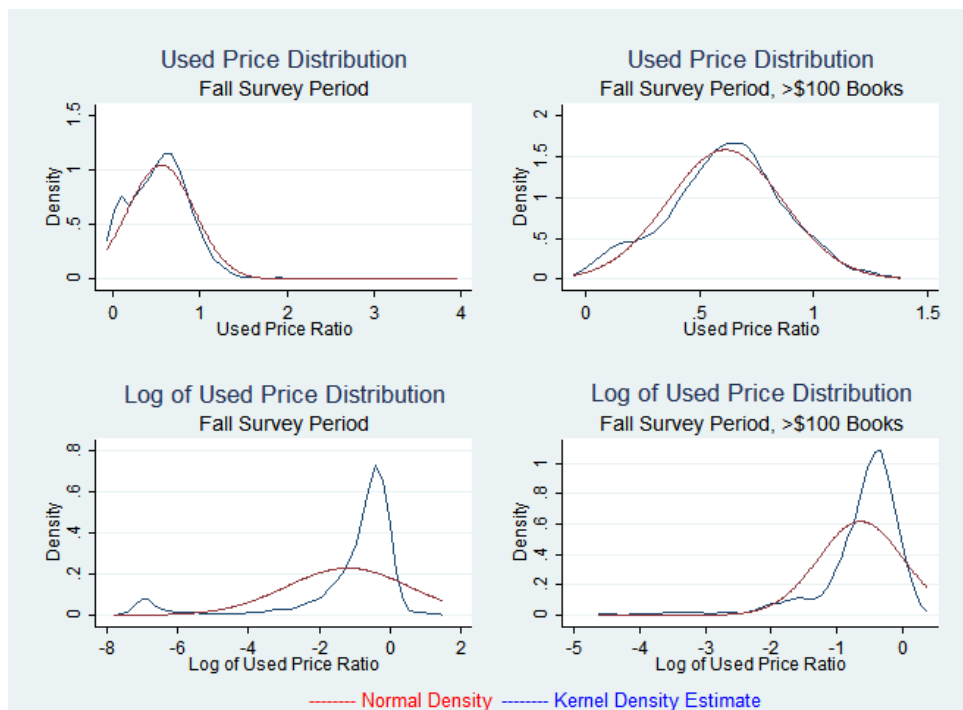


Figure A.2: Kernel Density Estimate

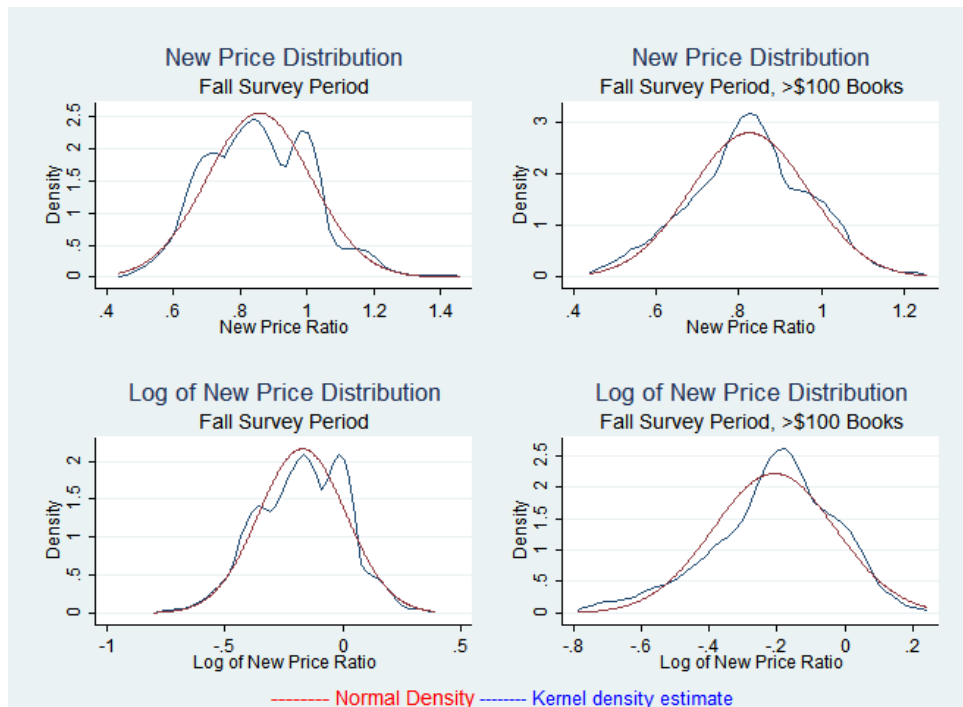


Figure A.3: Kernel Density Estimate

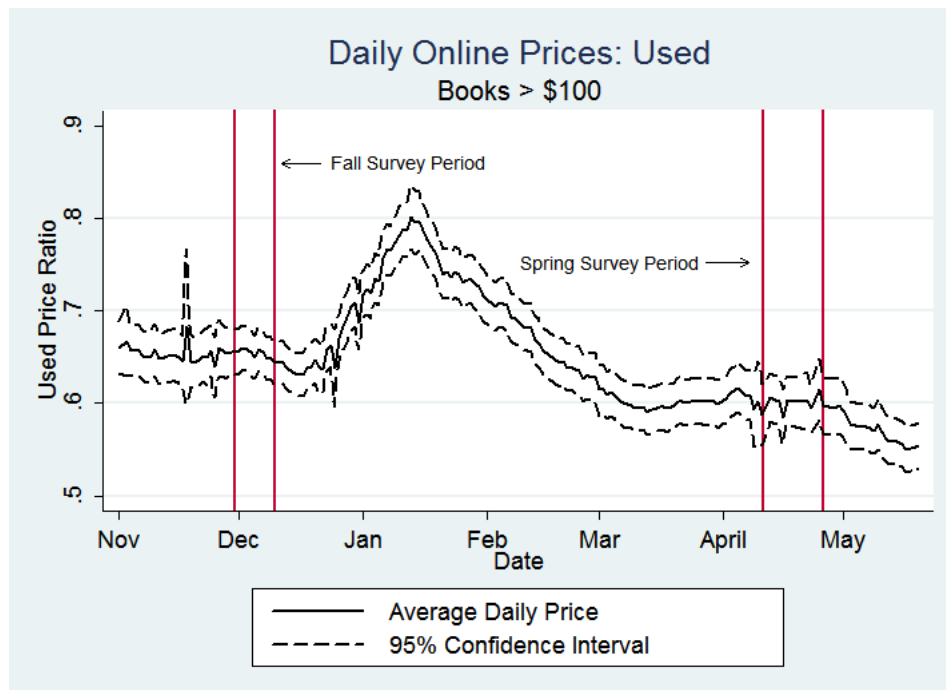


Figure A.4: Daily Used Prices

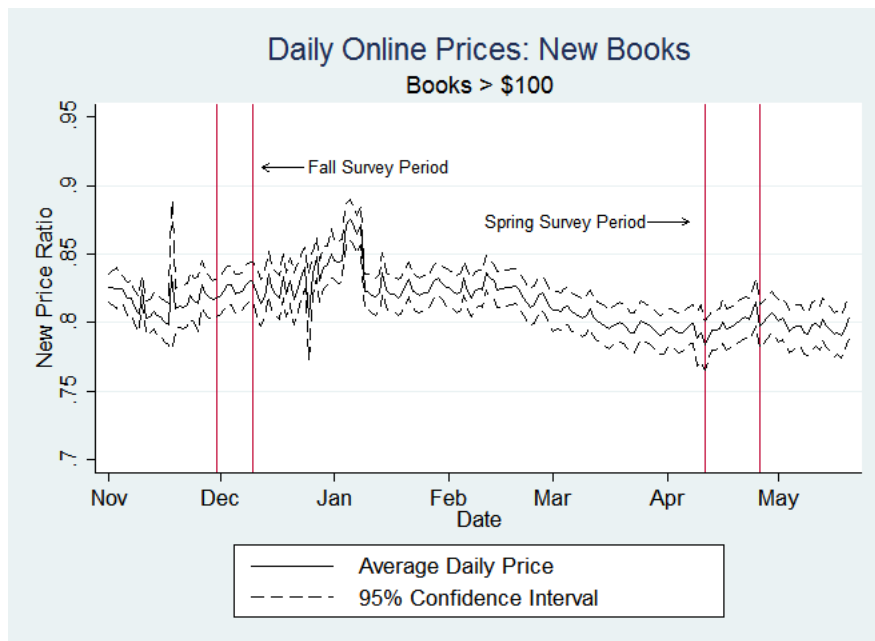


Figure A.5: Daily New Prices

BIBLIOGRAPHY

- Ackerberg, D. (2001). Empirically Distinguishing Informative and Prestige Effects of Advertising. *RAND Journal of Economics*, 32(2):316–333.
- Ackerberg, D. (2003). Advertising, Learning, and Consumer Choice in Experience Good Markets: an Empirical Examination. *International Economic Review*, 44(3):1007–1040.
- Arcidiacono, P., Hotz, J., and Kang, S. (2012). Modeling College Major Choice using Elicited Measures of Expectations and Counterfactuals. *Journal of Econometrics*, 166(1):3–16.
- Bresnahan, T., Stern, S., and Trajtenberg, M. (1997). Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980s. *RAND Journal of Economics*, 28(0):S17–S44.
- Bruno, H. and Vilcassim, W. (2008). Structural Demand Estimation with Varying Product Availability. *Marketing Science*, 33(2).
- Carlson, J. and Pescatrice, D. (1980). Persistent Price Distributions. *Journal of Economics and Business*, 33(1):21–27.
- Chevalier, J. and Goolsbee, A. (2009). Are Durable Goods Consumers Forward-Looking? Evidence from College Textbooks. *Quarterly Journal of Economics*, 124(4):1853–1884.
- Conlon, C. and Mortimer, J. (2012). Demand Estimation Under Incomplete Product Availability. *Working Paper*.
- Crawford, G. and Shum, M. (2005). Uncertainty and Learning in Pharmaceutical Demand. *Econometrica*, 173(4):1137–1173.
- De los Santos, B. (2008). Consumer Search on the Internet. *Working Paper*.
- De los Santos, B., Hortacsu, A., and Wildenbeest, M. (2012a). Search with Learning. *Working Paper*.
- De los Santos, B., Hortacsu, A., and Wildenbeest, M. (2012b). Testing Models of Consumer Search using Data on Web Browsing and Purchasing Behavior. *American Economic Review*, 102(6).
- Einav, L. (2005). Informational Asymmetries and Observational Learning in Search. *Journal of Risk and Uncertainty*, 30(3):241–259.
- Erdem, T. and Keane, M. (1996). Decision-making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets. *Marketing Science*, 15(1):1–20.
- Geweke, J. (1989). Bayesian Inference in Econometric Models Using Monte Carlo Integration. *Econometrica*, 57:1317–1339.

- Hajivassiliou, V. and McFadden, D. (1998). The Method of Simulated Scores for the Estimation of LDV Models. *Econometrica*, 66:863–896.
- Hong, G. and Shum, M. (2006). Using Price Distributions to Estimate Search Costs. *RAND Journal of Economics*, 37(2):257–275.
- Hortaçsu, A. and Syverson, C. (2004). Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds. *Quarterly Journal of Economics*, 119(2):403–456.
- Howard, J. and Sheth, J. (1969). *The Theory of Buyer Behavior*. John Wiley and Sons, New York.
- Keane, M. (1994). A Computationally Practical Simulation Estimator for Panel Data. *Econometrica*, 62:95–116.
- Koulayev, S. (2009). Estimating Search with Learning. *Net Institute Working Paper No. 08-29*.
- Koulayev, S. (2012). Estimating Demand in Search Markets: The Case of Online Hotel Bookings. *Federal Reserve Bank of Boston Working Paper No. 09-16*.
- Koulayev, S. (2013). Search with Dirichlet Priors: Estimation and Implications for Consumer Demand. *Journal of Business and Economic Statistics*.
- Marvel, H. (1978). Competition and Price Levels in the Retail Gasoline Market. *Review of Economics and Statistics*, 60(2):252–258.
- Matsumoto, B. and Spence, F. (2013). Price Beliefs and Experience: Do Consumers’ Beliefs Converge to Empirical Distributions with Repeated Purchases? *Working Paper*.
- Mehta, N., Rajiv, S., and Srinivasan, K. (2003). Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation. *Marketing Science*, 22(1):58–84.
- Moraga-González, J., Sándor, Z., and Wildenbeest, M. (2009). Consumer Search and Prices in the Automobile Market. *Working Paper*.
- Newell, A. and Simon, H. (1972). *Human Problem Solving*. Prentice Hall, Englewood Cliffs, NJ.
- Pires, T. (2013). Consideration Sets in Storable Goods Markets. *Working Paper*.
- Pratt, J., Wise, D., and Zeckhauser, R. (1979). Price Differences in almost Competitive Markets. *Quarterly Journal of Economics*, 93(2):189–211.
- Sonnemans, J. (1998). Strategies of Search. *Journal of Economic Behavior and Organization*, 35(3):309–332.
- Sorensen, A. (2001). An Empirical Model of Heterogeneous Consumer Search for Retail Prescription Drugs. *NBER Working Paper 8548*.

- Sovinsky Goeree, M. (2008). Limited Information and Advertising in the U.S. Personal Computer Industry. *Econometrica*, 76(5):1017–1074.
- Spence, F. (2013). Consumer Experience and the Value of Search in the Online Textbook Market. *Working Paper*.
- Stigler, G. (1961). The Economics of Information. *Journal of Political Economy*, 69(3):213–225.
- Stinebrickner, T. and Stinebrickner, R. (2011). Math or Science? Using Longitudinal Expectations Data to Examine the Process of Choosing a College Major. *NBER Working Papers 16869*.
- Train, K. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.
- Van Nierop, E., Bronnenberg, B., Paap, R., Wedel, M., and Franses, P. (2010). Retrieving Unobserved Consideration Sets from Household Panel Data. *Journal of Marketing Research*, XLVII:63–74.
- Wiswall, M. and Zafar, B. (2012). Determinants of College Major Choices: Identification from an Information Experiment. *Working Paper*.